Masters' Program in

Advanced experimental & computational Biosciences

Department of Biochemistry and Biotechnology

University of Thessaly

July, 2021

Analysis of Single Nucleotide Polymorphisms in regulatory elements of oncogenic lncRNAs

Laboratory of Molecular Biology and Genomics

Marios Lange

Three-member committee:

Supervisor

- Giakountis Antonios, Assistant Professor of Molecular Biology – Genomics

Members

- Mathiopoulos Kostas, Professor of Molecular Biology
- Karpouzas Dimitrios, Professor of Environmental Microbiology & Biotechnology

Table of Contents

Abstract

Gastric cancer is one of the leading forms of cancer both in regard to its prevalence and mortality, ranking fifth and second in the respective categories amongst other forms of cancer worldwide. This occurs due to its usually asymptomatic nature at its early stages and lack of suiting large-scale methodologies for prognosis and early detection for this form of cancer. This leads to late diagnosis, most frequently at a metastatic stage, where the applied therapeutical approaches are mostly ineffective. Therefore, novel biomarkers for the specialized prognosis and non-invasive diagnosis- including staging of the disease- of gastric cancer, are required for the personalized treatment of the patient. LncRNAs are involved in a huge variety of cellular processes as key regulators, and the fact that they have a cell-type and developmental/differentiation stage specific expression profile, makes them ideal focus points to identify candidate biomarkers for many forms of cancer, including gastric. Furthermore, identification and characterization of the effect of non-coding variants in regulatory elements of lncRNAs with a deregulated expression profile in gastric cancer, further clarifies the utility of the variant for a prognostic test and the lncRNA as a target in general. This study focuses on a candidate enhancer of RECUR1, an over-expressed lncRNA in gastric cancer, which carries an SNP correlated with increased chance of gastric cancer appearance and the survival probability of the patient. It was proven that the enhancer indeed has the potential to upregulate the expression of RECUR1, albeit further investigation is required. Additionally, the appropriate tools for the characterization of the role of the SNP were created and are ready to be used in an *in cellulo* approach.

1.Introduction

1.1. Gastric Cancer

1.1.1 Cancer types characterized by extreme variability: Gastric Cancer.

Gastric cancer is characterized by modest occurrence (5th rank world-wide, 3rd in Asia), with twice as many men suffering from the disease relative to women. The average age for diagnosis is 60 years old, while there is a rising number of cases reported between 40-50 years of age. The most affected countries are China, Korea, and Japan, due to genetic predisposition of the population in these areas, as well as due to dietary habits and exposure to xenobiotics (**Fig *1***) [1].

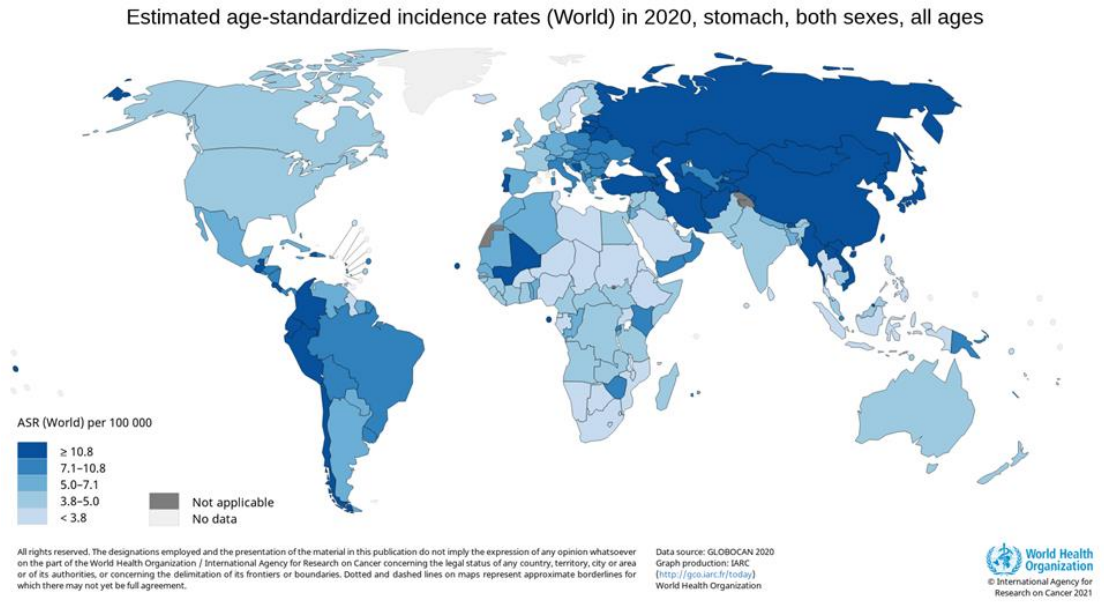

***Figure 1**: Incident rates of gastric cancer cases worldwide (Both sexes, all ages). Data from WHO/Cancer today 2020.*

Most importantly, though, gastric cancer presents the second highest patient mortality rates, being surpassed only by lung cancer (**Fig 2**). This is due to lack of early and routine diagnosis of the disease, which usually leads to detection at a later metastatic stage [2, 3]. As for the life-expectancy of the patient after the diagnosis, if the diagnosis occurs at an early stage, then 95% of the patients may be able to survive past the 5 years after diagnosis, while if the diagnosis occurs at a latent stage, then the life expectancy drops to less than a year. As for the therapeutic approaches, surgical excision of the tumor is considered obsolete in the advanced stages in which the diagnosis takes place, and not many sufficient therapeutic approaches have been found [4, 5]. Thus, new molecular targets are needed both to ensure early diagnosis of the disease, and to create novel treatments for each individual patient [6, 7].

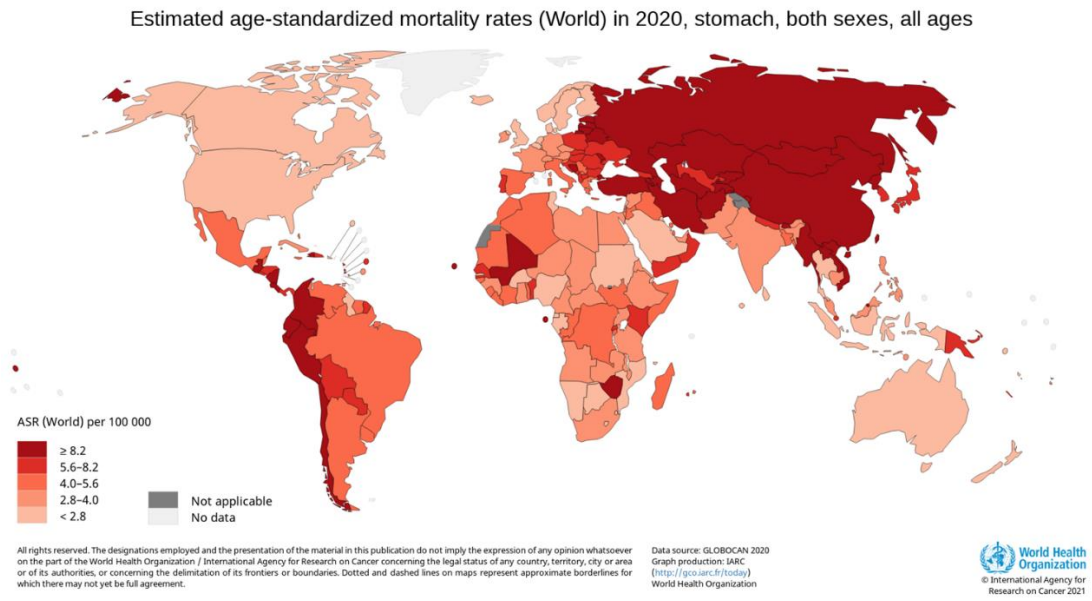

Figure 2: *Mortality rates of gastric cancer cases worldwide (Both sexes, all ages). Data from WHO/Cancer today 2020.*

1.1.2 Phenotypical characteristics of gastric cancer cells

The histological characteristics of gastric cancer, which include a classification in regard to the morphology of the tumor, are well characterized and used for the detection of the tumor via gastroscopy. More specifically epithelial cells are most commonly transformed leading to appearance of adenocarcinoma (95% of the total cases), followed by the less frequent mesenchymal, lymphoproliferative, and neuroendocrine types [8]. Further classification of adenocarcinomas is possible, with three prominent types [9]:

1. Intestinal type, in which the morphology of the tumor is similar to that of the intestinal tube, in most cases due to existing intestinal metaplasia. It is also characterized by cylindrical, gland-like form and thin stroma. This type of cancer appears at 54% of the cases, and affects older patients, mostly males.
2. Diffused type, which is characterized by high EMT, poor cell cohesion and thick stroma, while it is also mutation prone. It appears in 32% of adenocarcinoma cases, and it is heavily linked with appearance in younger age.
3. Mixed type, which shares common characteristics with both types.

1.1.3 Genetic characteristics of gastric cancer cells

Gastric cancer carries a heavy mutational burden, that either preexists in the genome in the form of inherited polymorphisms (SNPs), or by somatic mutations and alterations occurring by xenobiotic factors or infectious diseases (such as infection by Epstein-Barr virus or *Helicobacter pillory*). These lead to CIN (which affects large areas of the genome) or to Microsatellite instability (MSI-in which the lesions are restricted to a smaller scale), or to genomically stable molecular subtypes. According to The Cancer Genome Atlas (TCGA) consortium, gastric cancer cases can be divided into 4 distinct molecular subtypes based on their genomic signature [10-14]:

1. Tumors that correlated with *Epstein-Barr* virus infection (EBV+) and appear in 8,8% of the patients. This subtype is characterized by amplification of 9p24.1 chromosomic area (carries the JAK2 and PD-L1/2 genes), mutations in various oncogenes or tumor suppressors such as PI3K, ARID1A and BCOR, while it also shows silencing of CDKN2A and rarely mutated p53. Additionally, it shows alterations in methylation profile and infiltration of the tumor by T-cells.
2. Tumors with MSI, that appear in 21,7% of the patients. This subtype is an aggressive form that appears in old age, is non-responsive to adjuvant chemotherapy and is characterized by hypermutations in various genes, such as TP53, PI3KCA, ERBB3 and ARID1A and silencing of MLH1 and various others.
3. Genomically stable tumors, that appear in 19,7% of the patients. This subtype carries multiple mutations in genes relative to cell adhesion such as CDH1 and RHOA, fusion of CLDN18 with ARHGAP and enrichment in pathways related with angiogenesis. This type is correlated with diffuse histology and appearance at a young age.
4. Tumors with CIN, that take up 49,8% of the total cases. This subtype is mostly correlated with intestinal histology metaplasia, and shows extensive mutations in TP53, SMAD and APC, as well as activation of the RTK-RAS pathway, along with molecules such as EGFR, JAK2 and VEGFA.

In each of these subtypes, various SNPs are involved with either the tumorigenic process or maintenance of gastric cancer cells, depending on their location in the genome. These SNPs can have functional impact to a variety of protein coding genes, such as the DNA damage repair machinery, xenobiotic metabolism, molecule biogenesis processes, or other gene products that are involved in carcinogenesis, such as tumor suppressors p53 and CDH1 [15-19]. A special category of SNPs involved in tumorigenesis in the case of gastric cancer, includes SNPs in genes that regulate the defense against *Helicobacter pylori* infections [20, 21]. Some SNPs that are positively correlated with the appearance of gastric cancer are the intronic SNP rs2976392 (A>G) in the *PSCA*(prostate stem-cell antigen) and intronic SNP rs13361707 (T>C) in *PRKAA1*(which encodes Protein Kinase AMP-Activated Catalytic Subunit Alpha 1) [22, 23].

Chromosomal instability is caused by presence of lengthy copy number variations (CNVs) or by chromosomal translocations and is one of the main characteristics of gastric cancer. Furthermore, some chromosomal areas are susceptible to these alterations in different types of gastric cancer, with alterations in 8q, 17q and 20q being prominent in the intestinal subtype of gastric cancer, while in the diffused type 12q and 13q are the top affected chromosomal areas [24-26]. Due to these, multiple reports have shown various loss of heterozygote events (*TP53, APC*), gene fusions such as the one of SLC1A2–CD44 which appears rarely in cancer cells but is a main driver of their disorganization and enhancer hijacking as in the case of HER2 (tyrosine kinase receptor, part of the MAPK signaling pathway) in 20% of the patients [27-29].

The extent of microsatellite instability differs between tumors, classifying them into tumors with low and high presence of MSI, depending on the extent of mutation occurrence, with alterations in microsatellite length not having a distinct role yet [30]. This phenomenon occurs due to deregulation of enzymes that take part in the repair mechanisms of the cell, and most prominently in the base-excision repair (BER) mechanism due to epigenetic alterations at their promoters. This leads to extended presence of mutations in the genome, especially in areas with repeated sequence motifs (hence the mutational burden at microsatellites), leading to deregulation of many cellular processes [31].

As with every other type of cancer, the majority of the mutations occur in non-coding parts of the genome. Some examples include:

- Introduction of cryptic promoters in areas that in most cases exceeded the 500bp distance from the TSS, leading to non-canonical mRNA transcripts [32].
- Mutations at CTCF binding sites linking to CIN, due to the disruption of the CTCF Binding Site, leading to altered expression of the neighboring genes (see also below). In some cases, such alterations occur in a tissue-specific manner, as specific mutations typically predispose only for gastric (and colorectal in some extent) cancer onset and rarely for other types [33].
- Alterations in super-enhancer sequence, epigenetic profile and transcription factor occupancy, due to mutations occurring in the area and their enrichment in SNPs and copy number alterations [34].
- Alterations of lncRNA secondary structures, that modify regulatory interactions formed by the lncRNAs with RNA-binding proteins and/or miRNAs [35].

Thus, further combined studies of non-coding variants in regulatory elements and ncRNAs has a pivotal role in finding new targets for prognosis, diagnosis, and personalized treatment of patients with gastric cancer.

1.2 The human genome.

1.2.1 Exploring the human genome

The human genome has a complex structure, with its functional organization being heavily regulated both in terms of structural conformation and DNA sequence composition. Within a chromosome, a certain level of compartmentalization can occur, with prominent examples being: 1) lamina associated domains (LADs) being formed near the nucleolar lamina of the inner nuclear membrane by heterochromatinic regions, and 2) topologically associated domains (TADs) whose boundaries are restricted by presence of the CCCTC binding factor domains (CTCF motifs). Within the TADs, the formation of enhancer-promoter loops is enabled, which in turn regulate gene expression (***Fig.3***)[36-38].

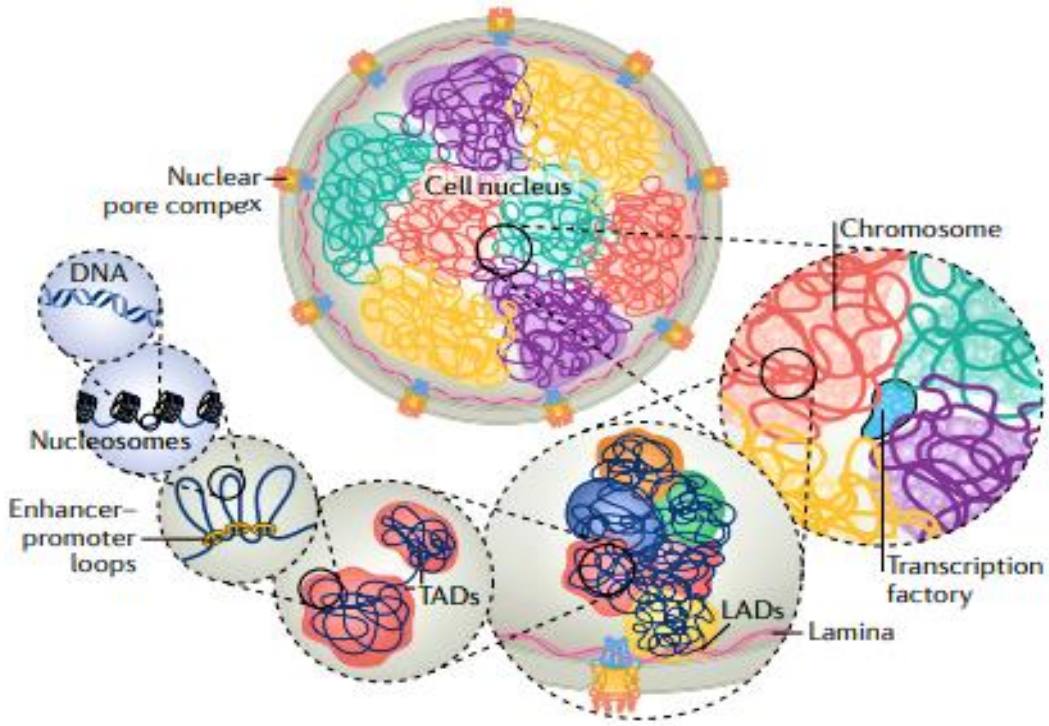

***Figure 3**: Structural organization of the DNA within the nucleus.*

Since the publishing of the first human genome sequences by the Human Genome Project, approximately 24,000 protein coding genes have been found (1.5% of the genome and ~2% of the transcriptome), while the inclusion of non-coding RNA genes (1.5% of the genome and ~98% of the transcriptome) has raised the total number of genes to around 40,000, with this number still being debatable since it does not account for gene isoforms, that extend beyond 100,000 [39, 40]. Part of the genome hosts various regulatory elements such as: 1) enhancers, which are sequences that carry transcription factor motifs that induce gene expression 2) silencers, that carry motifs for repressive transcription factors and 3) insulator sites that regulate the architecture of specific loci, such as CTCF binging sites that determine the limits of TADs, (see also below) [41-43]. Each element is characterized by its own set of histone modifications, and transcription factor binding motifs. The communication between promoters and enhancers is enabled by loop formation within a TAD, the stabilization of which can be regulated by lncRNAs that ultimately enable or block gene expression (***Fig.4***) [44, 45].

The rest of the genome mostly comprises of repetitive DNA sequences like LINEs and SINEs, as well as various others whose role is yet to be characterized. At the population level, the majority of the DNA sequence is identical to the consensus received by HapMap project and lately the Genome Reference Consortium (GRCh38.p13), yet single nucleotide polymorphisms are responsible for 0.1% variation between individuals, that raises up to 0.6% percent when we take indels into account [46-49].

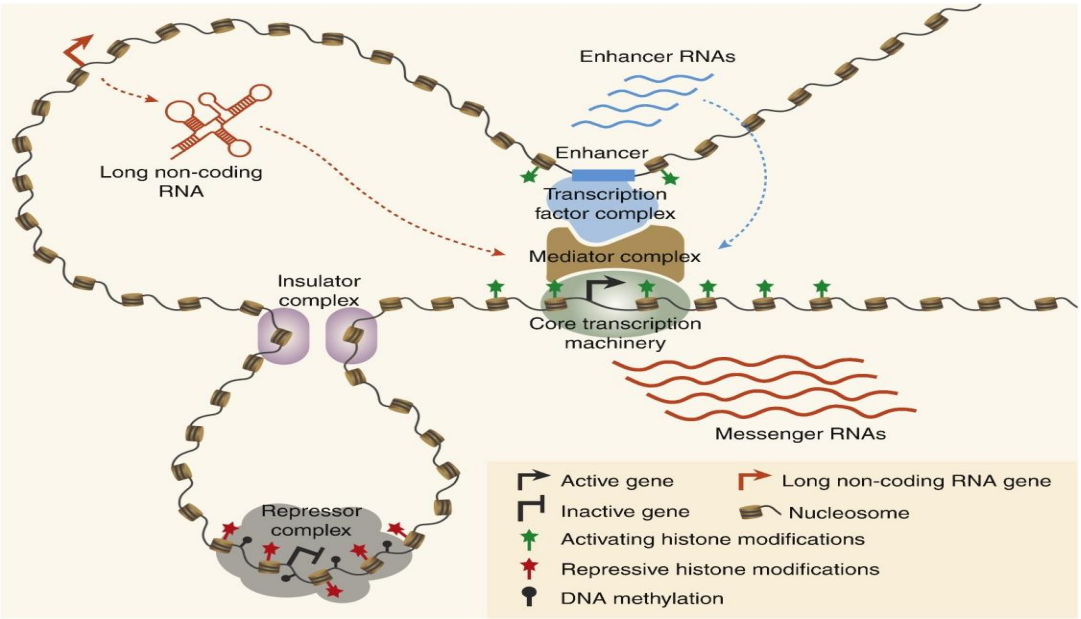

***Figure 4**: The role of DNA regulatory elements and lncRNAs in gene expression.*

1.2.2 The cancer genome.

Naturally occurring variation, in the form of Single Nucleotide Polymorphisms (SNPs) can predispose for tumorigenesis, since a portion of them has been associated with tumor initiation, progression, and metastasis [50-53] through genome-wide association studies (GWAS). Therefore, SNP alleles indicative of a disease or cancer specific subtype serve as risk factors, depending on the strength of the allele that they carry. However, due to the presence of linkage disequilibrium in the genome, extensive research needs to be conducted in order to distinguish the true positive driver alleles, from the neutral and co-segregated passenger mutations [54-56].

On top of naturally occurring SNPs, the cancer genome gains extensive variability, through mutations due to exposure to xenobiotic factors with mutagenic potential, a persons' lifestyle (e.g., smoking) or UV radiation. These mutations may appear in every part of the genome, either by causing single nucleotide variations (SNVs), in the form of insertions, deletions or duplications, or they may affect wider areas (normally above 50bp and up to 1Mb), in which we have aneuploidy or duplications of the target sequences, thus called copy number variations (CNVs) [57-59]. CNVs may occur either in germline cells, thus being correlated with hereditary types of cancer (e.g., familiar breast cancer), or they occur in somatic cells, thus being present in various cancer types, especially those characterized by Chromosomal INstability (CIN). CNVs can lead to carcinogenesis through 3 distinct mechanisms: 1) gene dosage alterations, which may affect at least one of the two copies of the gene, 2) gene fusions, and 3) alterations occurring on the genes' regulatory elements, which may in turn affect the nature of the *in cis*- or *in trans*- interactions formed between the regulatory elements and the promoter of the gene-target [60, 61]. In rare occasions, presence of CNV has been correlated with alterations in the methylation state of regulatory elements, both due to introduction or loss of CpG islets [62]. Such mutations (SNVs and CNVs) have already been shown to cause alterations in proteins through non-synonymous amino-

acid changes that may affect the structural integrity and/or function of the protein in various degrees , premature ending of its translation through creation of a stop codon in a protein coding gene, altering of the splicing profile through gain or loss of a splicing site, gene fusions both in coding and non-coding genes, duplications or deletions of genes or their regulatory elements and sequences within them (e.g., promoters, enhancers), or even alterations in insulation profile, that may result in ectopic chromatin architecture [63-67].

The vast majority of the genome and thus the areas in which mutations may occur, comprises of genomic areas outside the coding sequences of the protein-coding genes. Thus, additional studies are required to shed light to the role of these mutations in non-coding regulatory elements and in non-coding RNA genes, the downstream function of which is yet to be characterized. More specifically, a possible classification of these non-coding mutations according to the genomic area that they occur, refer to mutations that occur within [68, 69]:

- Regulatory elements that lie within the transcribed part of a gene (introns, 5' and 3' untranslated regions-UTRs) which can be found both in protein-coding and non-coding transcripts, and
- Regulatory elements that exceed the transcribed area of the gene (promoters, enhancers-with the exception of enhancers that lie within intronic sites-silencers and insulators).

Of special interest are mutations that target non-coding RNAs, as their transcript often has a regulatory role, thus can be analyzed further as a sub-category of mutations in RNA molecules with a regulatory function [70].

1.2.2.1 Mutations in transcribed regulatory elements.

Mutations that target transcribed regulatory elements have various effects in regard to the processing, translation (in protein-coding transcripts) and stability of the affected transcripts, while in non-coding RNAs it may also affect their ability to execute their regulatory function [67, 71].

1.2.2.1.1 Mutations in intronic regions

Intronic areas can be found in many eukaryotic genes. They have many important roles, as at the DNA level they may contain enhancers and/or silencers, and at transcript level they contain the regulatory elements for splicing and alternate splicing. More rarely, introns may also have a role in mRNA exportation from the nucleus, nonsense mediated decay (along with the UTRs), or could even be used to produce ncRNAs [72-74]. On some occasions, the intron-mediated enhancement (IME) phenomenon may occur, in which the expression levels of the gene that contains the intron may be elevated with the effect being present only when the intronic sequence is transcribed [75]. Another important function of introns is to protect the cells from transcription-mediated genome instability by preventing the formation of R-loops, thus blocking the recombination activities that may occur in this state. This function is extensively found in highly expressed genes [76]. Many diseases and cancer

cases are correlated with splicing alterations, which either lead to exon skipping or intron retention, or are caused by mutations that create new splicing sites, thus altering a genes' splicing profile [77]. Occurrence of SNVs in the vicinity of splicing sites and exon-intron junctions may result in altered splicing profile, which in some cases of tumor suppressor genes such as *POT1* (telomere maintenance gene) lead to inclusion of exons containing a premature termination codon in blood cancer [78]. Further study of this category is needed, as splicing alterations are major players in each "hallmark" cancer process that give the tumor a more aggressive identity, pinpointing alternative splicing inhibitors as possible therapeutic solutions in cancer [79].

1.2.2.1.2 Mutations in 5' and 3' UTRs

Both the 5' and 3' UTRs have important roles to play in the post-transcriptional modifications and translation of the transcript. The 5' UTR lies upstream of the coding sequence and carries the ribosomal binding site, internal ribosome entry sites, as well as upstream open reading frames (uORFs). The 3' UTR lies beyond the stop codon of the coding sequence and carries elements for the regulation of translational termination, transcript modification and miRNA target sites. Many mutations have been found in the UTRs correlating to the general abundance of the transcripts of genes related to carcinogenesis, such as Progesterone receptor (PGR) and Erb-B2 Receptor Tyrosine Kinase 4 (ERBB4) [80]. An example of a SNP within a UTR that is linked to cancer predisposition, is the *POMC* c.*28delT (rs756770132), which lies in the 3'UTR of the gene encoding pro-opiomelanocortin. The product of this gene acts as a precursor for many signaling molecules, thus linking it to energy metabolism of the cancer cells and immune reactions, while alterations of its expression levels are linked to increased DNA damage [81]. Recurrent mutations and SNPs in *NFKBIZ* 3'UTR have been correlated with activation of the NF-κB signaling pathway in diffuse large B-cell lymphoma, due to altered transcriptional regulation and increased transcript levels [82]. A graphical abstract of the effects, along with some examples of genes affected by mutations in the UTRs can be found on ***Figure 5*** [83].

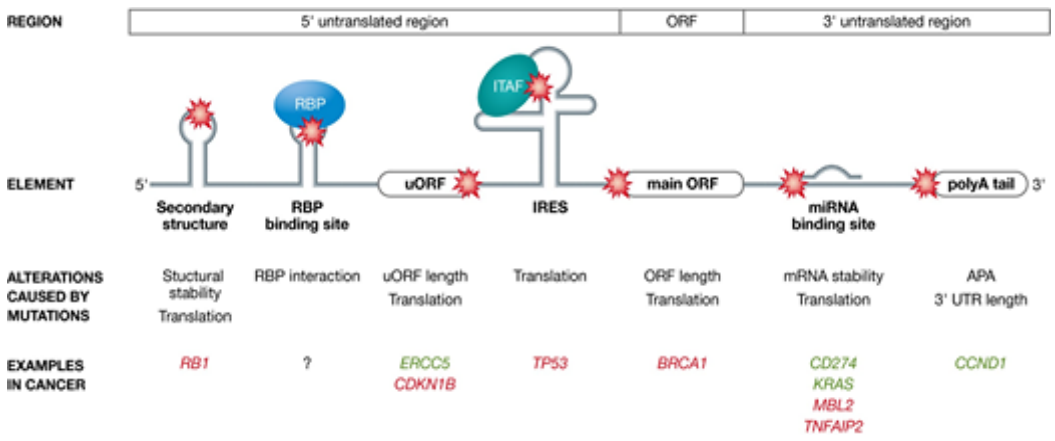

Figure 5*: Mutations altering the secondary structure of the mRNAs 5'- and 3'-UTRs. These mutations may occur in the secondary structures formed to increase the structural stability of the transcript, interaction with transcription factors (RBP binding site, IRES) or upstream ORFs in the 5'-UTR, while they may affect the poly-adenylation site or miRNA target sites on the 3'UTR. In tumorigenesis, the*

transcripts of oncogenes tend to have induced stability or alternative ORFs leading to their oncogenic effect (gene transcript examples in green coloration), while transcripts of tumor-suppressors tend to have shorter stability, translation capacity and increased miRNA targeting (examples in red coloring).

1.2.2.2 Mutations in non-transcribed regulatory elements.

Many non-coding variants have been identified within non-transcribed regulatory elements and are heavily linked with abnormalities in regulation of transcription. These changes may directly affect the function of the specific element at their genomic area within a TAD or may occur due to changes that affect the architectural organization of a specific locus, thus enabling interactions of elements that would not interact under normal conditions. The key interaction that is affected, is the formation of promoter-enhancer loops and the functionality of these elements *per se* [84-86].

1.2.2.2.1 Mutations in promoter elements

Focusing on promoters, and more specifically on the core promoter area, these are the main drivers of transcription, as they carry motifs for general transcription factors required for the attraction of the RNA-polymerase near the transcriptional start site (TSS). In general, promoters can be characterized by the presence of a plethora of transcription factor motifs, their (usually) directional activity and ability to activate basal levels of transcription [87, 88]. Also, they carry distinct epigenetic signatures for active, poised, bivalent and inactive states each characterized by the different methylation state of Lysine 4 in histone 3, and the presence of acetylation or methylation of Lysine 27 in Histone 3 [89-91]. Promoter mutations in the TERT promoter are prominent in various cancer forms (bladder, urothelial, melanoma etc.), and especially occurrence of 2 specific SNVs, the -124 C>T and -146 C>T, which lead to transcription of the TERT gene (***Fig. 6***) [92, 93]. The effects of an SNP within a promoter comprise of creation of a cryptic promoter/enhancer, alteration of transcription factor binding motifs (that may lead to the disruption of the promoter, or creation of a motif with stronger binding affinity to the transcription factor) and differential methylation of the promoter due to loss or gain of CpG islets [94-98]. Such an example is the case of the dominant G allele of the rs11672691 (G/A) SNP, which lies within the promoter of the lncRNA PCAT19, turning the promoter into a cryptic enhancer that positively regulates the expression of the long isoform of PCAT19, which is linked with aggressive form of prostate cancer [99, 100]. Many promoters also carry binding motifs for the repressive Yin-Yang 1 (YY1) protein, and it has been shown that presence of the high-risk C allele of the SNP rs17079281 (C/T within the promoter of *DCBLD1,* leads to a disruption of the binding motif of YY1, thus removing the repressive effect of the TF to the oncogene and initiating lung cancer in European and Asian populations[101]. Another prominent example of promoter affecting mutations, is the occurrence of either germline or somatic CNVs in the promoter of BRCA1, leading to the destruction of the promoter of the tumor-suppressor gene and carcinogenesis based on the two-hit model[102]. The opposite approach, which refers to activation of oncogenes, applies to various subtypes of lung adenocarcinoma, in which destruction of methylation sites at promoters has been found [103].

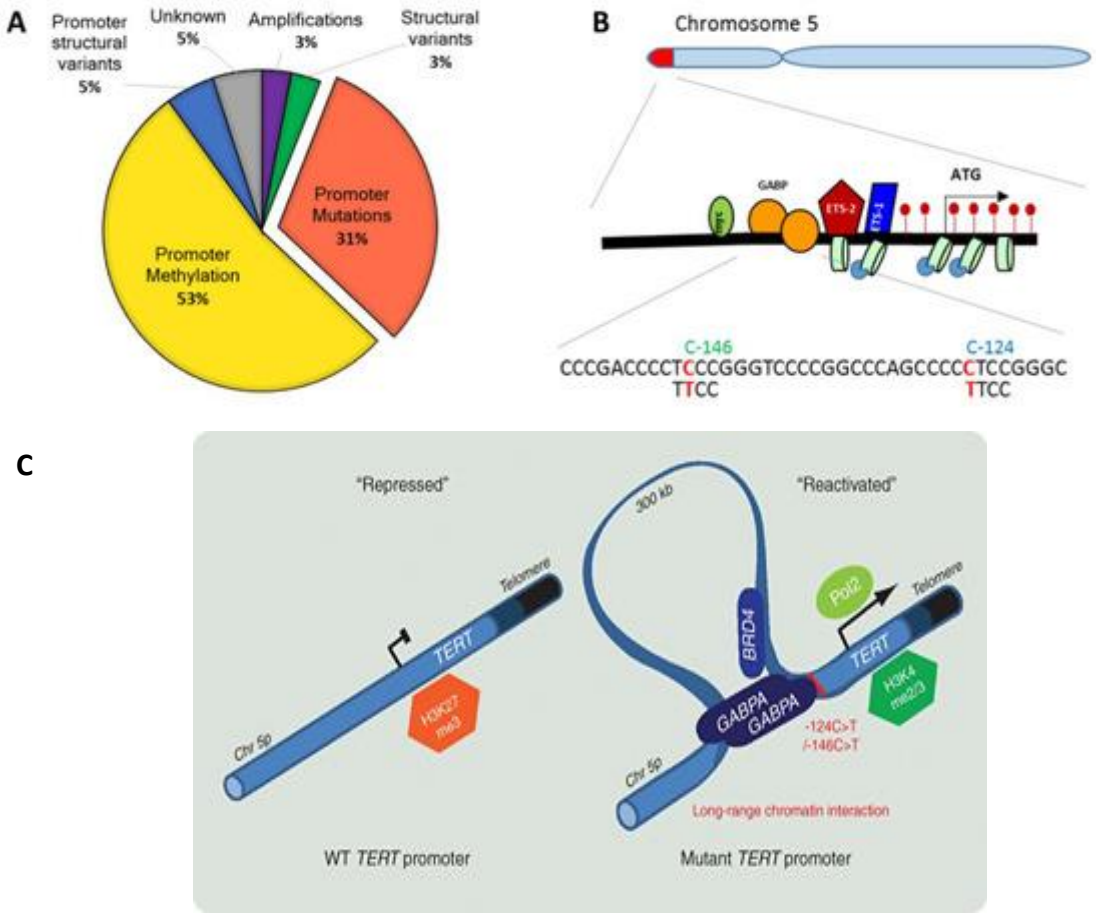

***Figure 6**: Mutations in TERT promoter. A) Chart pie of the different alterations occurring in the TERT promoter during its reactivation in cancer. B) Localization of the -124 C>T and -146 C>T mutations of the TERT promoter. C) The 2 mutations promote looping of the promoter with a putative enhancer region, leading to reactivation of TERT expression*.

1.2.2.2.2 Mutations in enhancer elements

Genomic areas that carry binding motifs for activating transcription factors and interact with promoters usually within a TAD in the 3D space, are characterized as putative enhancers [104-107]. Enhancers are also characterized by developmental stage and cell-type specific manner of activation, as not all enhancers are active in a specific cell-type in each given time of its development, while some enhancers may also act as silencers in some cell types due to the duality through which their associated transcription factors act (prime example are the enhancers that bind CDX2, which frequently have repressive effects in the cancerous intestine and activating role in gastric cancer) [108, 109]. On an epigenetic level, several histone modifications that decorate enhancers, mark them as active, poised, and inactive, depending on the presence of H3K4me1, H3K4me2, H3K4me3, H3K27ac and H3K27me3 histone modifications (***Fig.7***) [110].

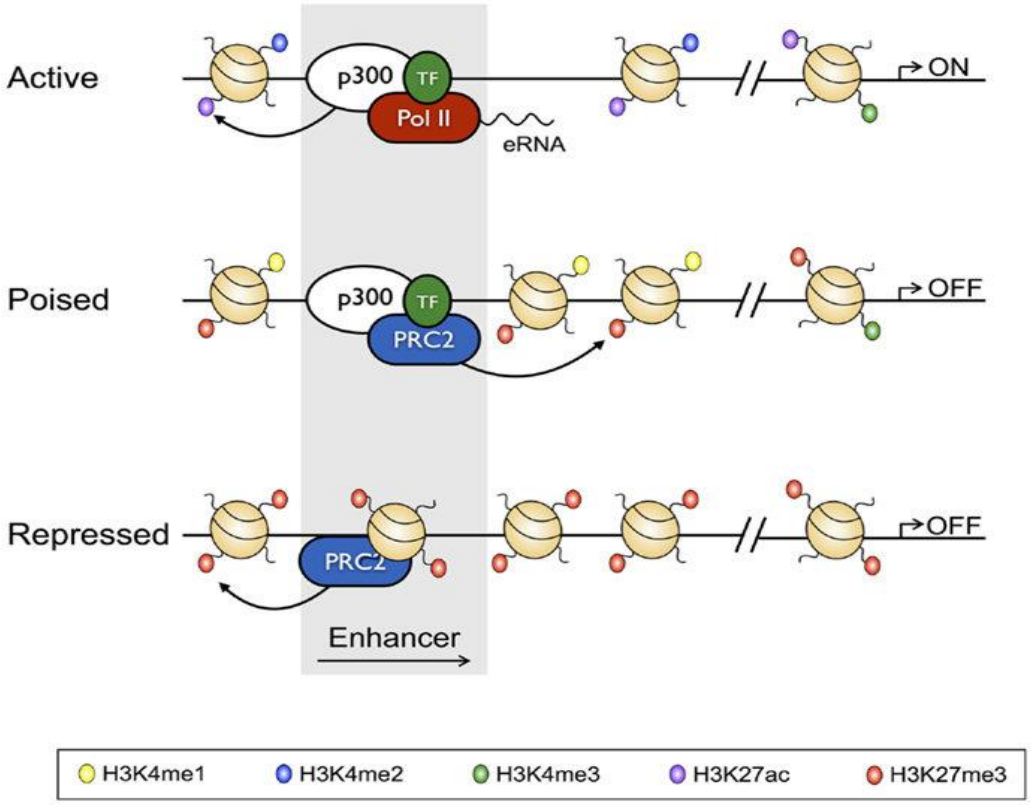

***Figure 7**: Enhancer activity based on their histone modification signature. Active enhancers are characterized by the presence of H3K4me2 and H3K27ac, while poised enhancers show increased levels of H3K4me1 and H3K27me3. Finally, repressed enhancers show extended presence of the H3K4me3 histone modification.*

Since the enhancers may act independently of their location in relation to the promoter they activate, they can be found anywhere within the corresponding TAD, while loop formation, ensures close proximity with the target promoter in the 3D space [111]. Enhancers can be also characterized by expression of enhancer RNAS (eRNAs), which are short, bidirectional transcripts produced at the site of the enhancer, either prior to or following the formation of the enhancer-promoter loop, depending on the case, with a suggested role for these transcripts being the stabilization of this loop along the formation of the mediator complex, in order to proceed to the expression of the target gene [112-114]. When multiple enhancers that show high enrichment in transcription factor motifs and extended transcription of eRNAs at their genomic sites, aggregate on a genomic area, they are deemed as Super-enhancers. Super-enhancers span wide genomic areas and can regulate the expression levels of pluripotency genes in stem cells, as well as other genes in cell-type specific manner. A summary of the above can be seen in ***Figure 8*** [115, 116].

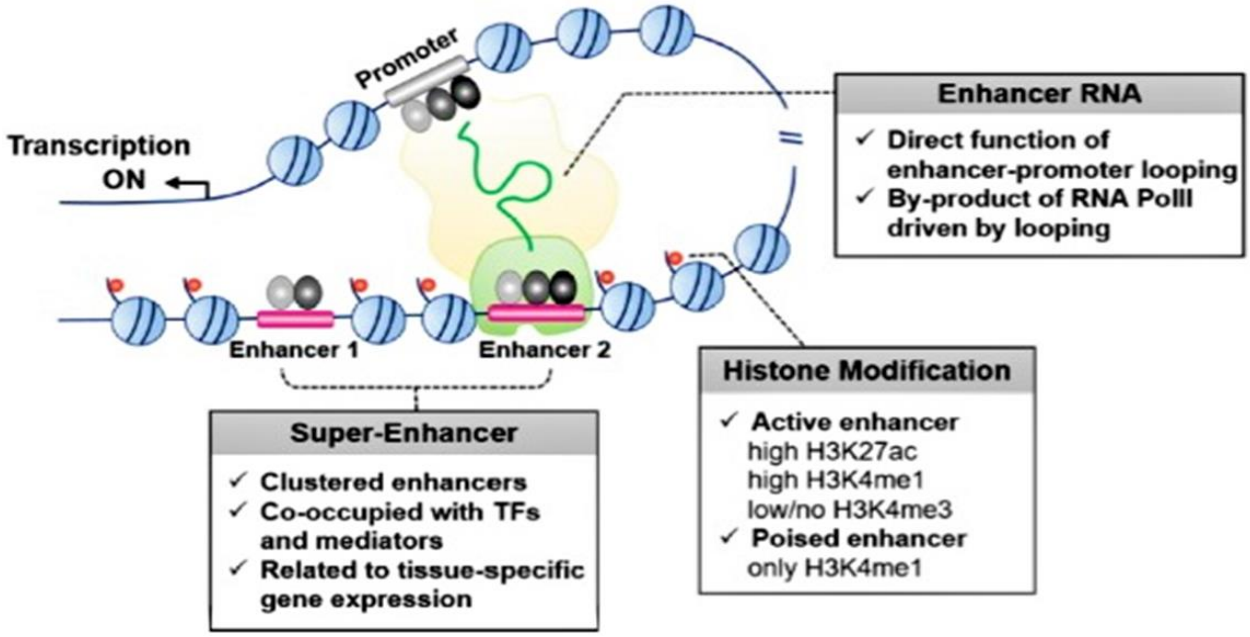

***Figure 8**: Enhancer activity in promoting gene expression. Bothe enhancers and super-enhancers may form loops with promoters, each of them occupied with the transcription factors able to regulate the gene expression. Stabilization of the enhancer-promoter loop is performed by the eRNAs, which are bi-directionally transcribed at the enhancer elements by RNA-polII. The enhancer may be either in an active or poised state, which is regulated by the histone modification signatures around the enhancer.*

Ectopic activity of enhancers and super-enhancers is heavily linked with disease onset and progression, due to their inherent role in upregulating a gene's expression levels. Thus, deregulation of these elements, either due to variants that alter transcription factor binding affinity, variants that disrupt or lead to the creation of (super)enhancers or variants that alter TAD boundaries, giving these elements new promoters to interact with, is crucial for tumorigenesis (***Fig.9***) [117].

Such an example is the SNP rs11672691 in the promoter of lncRNA PCAT19, which is found within the binding motif of the HOX2A TF in an enhancer regulating the expression of the lncRNA. The G allele of this SNP increases binding affinity of the transcription factor, leading to stabilization of the enhancer-promoter loop and increased expression levels of the oncogenic lncRNA [118]. Another example of an enhancer that carries an SNP that alters the binding capacity of TFs, is an enhancer regulating *MYC* and the lncRNA *PVT1*. The allele AC of the SNP rs35252396 variant (AC>CG), is positively correlated with increased binding of the hypoxia inducible factors (HIFs) to this enhancer, thus promoting survival of the cancer cells, with a prominent role in renal adenocarcinoma [119]. Moreover, CNVs that cause creation of super-enhancers are considered hallmarks of cancer, with a prominent regulatory effect in various driver oncogenes in many different types of cancer [34].

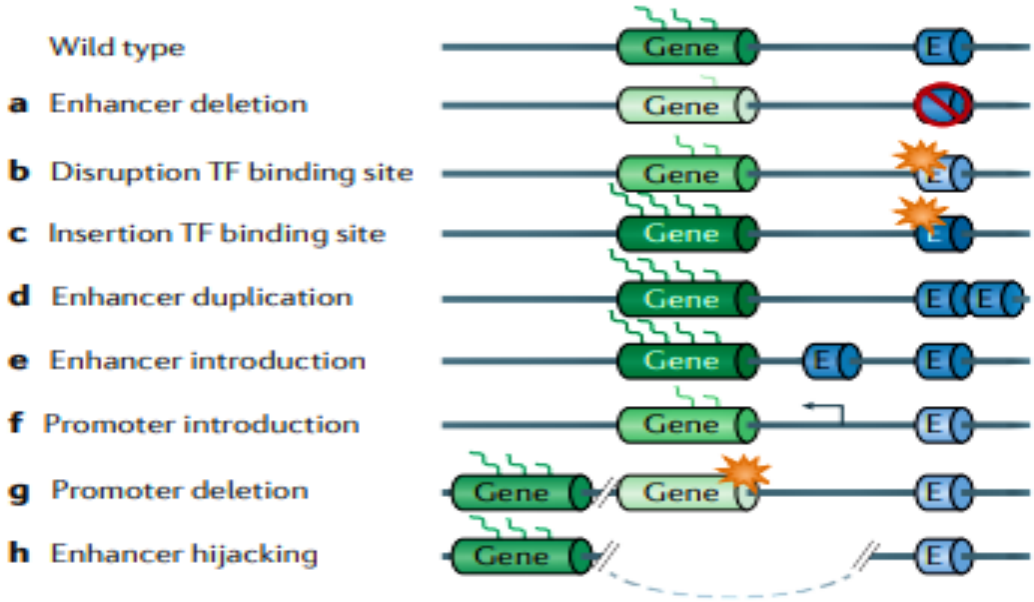

***Figure 9**: Mutations affecting the enhancers' regulatory wiring to target-promoters. Enhancer deletion due to presence of CNVs(a), or due to a point mutation or SNP that disrupts a binding motif of a TF crucial for its function(b). An SNP or point mutation may also create a binding motif with stronger affinity for a TF, or a CNV might duplicate part of the enhancer (c), or the entire enhancer (d) if they are found within the enhancer, or they can create a de novo enhancer (e). Similarly, they can create a new promoter for the enhancer to activate (f) or delete an already existing one, thus the enhancer might activate the next closest promoter within a TAD (g). Finally, when the TAD boundaries are compromised, then the enhancer may activate a gene promoter which lies at a different TAD, that they would never interact under normal circumstances, with the phenomenon being described as enhancer hijacking(h).*

1.2.2.2.3 Mutations in silencer and insulator elements

Studies of silencer and insulator elements are relatively new, with even fewer data available with regards to the role of specific variants that lie within them. Both elements have a significant fine-tuning role in a gene's regulation of expression [120]. Silencers carry binding motifs for repressive transcription factors, or histone methyltransferases, that catalyze the repressive trimethylation of lysine 72 of histone 3 in the promoters of the gene targets [121]. Some silencers have the ability to bind transcription factors both in their unmethylated and lowly methylated forms, such as FRA1, BACH2 and EBF1 transcription factors [122-124]. An example of a SNP that destroys a binding motif for a repressive TF can be the SNP rs249473, which lies within a silencer element, with the A high risk allele disrupting the binding motif for YY1. Under normal circumstances, YY1 would act in this case by downregulating the expression of AKT1, which is part of the PI3K/AKT/mTOR pathway, a pathway whose deregulation has a prominent role in various types of cancer, such endometrial and thyroid cancers [125]. As for the insulator elements, they are rich in binding motifs for CTCF, and their main function is to create interchromosomal barriers that either insulate different TADs, or act as a limit for heterochromatic regions[126]. Loss of insulator elements mostly due to CNVs has been described in various CIN cancers with these elements acting as the main drivers for the enhancer hijacking phenomena [127].

1.2.2.3 Mutations in non-coding RNAs.

Non-coding RNAs (ncRNAs) have a plethora of regulatory roles, and are directly involved in many cellular processes, such as regulation of the cells' growth rate and division, differentiation and apoptosis. Also, many ncRNAs show cell- and stage-specific activation pattern, further signifying regulatory fine tuning as their primary mode of action. Due to this fact, their deregulation has a prominent role in cancer onset, development, rate of differentiation and tumor metastasis [128-132].

Depending on their length, ncRNAs can be distinguished into short non-coding RNAs such as miRNAs, with a processed transcript length that is less than 200nt, and into long non-coding RNAs (lncRNAs), the transcript lengths of which exceed 200nt [133]. Key regulators for the different stages of the disease fall in both categories, so their study can be beneficial both in finding new prognostic and diagnostic markers, but also possible therapeutical targets, with higher specificity than the existing ones, which mainly refer to mRNAs [134, 135].

1.2.2.3.1 Role of miRNAs and their mutations.

miRNAs are small ncRNAs that are transcribed by MIR genes, and their main function is to target mRNA transcripts, inhibiting their translation and/or leading them to degradation [136, 137]. Depending on their effect in cancer, miRNAs can be classified as oncogenic miRNAs (oncomiRs) and tumor-suppressive miRNAs(tsmiRs), thus their expression profile can be useful for monitoring the disease state. Their value as diagnostic markers is also highlighted by their ability to maintain their integrity in various biological samples, with most prominent being blood serum [138, 139]. OncomiRs have increased expression levels in cancer cells and are linked with increased ability of the cancer cells to survive, to execute the epithelial to mesenchymal transition (EMT), and to increase the metastatic ability of the cells, through the degradation of tumor suppressor genes [140]. On the other hand, the transcriptional levels of tumor suppressing miRs are reduced in cancer cells, as their function is to target oncogenic mRNAs, thus mutations that lead to downregulation or loss of these transcripts and their subsequent genomic loci provide an advantage for the survival of the cancer cell [141]. A single mature miRNA product can have multiple targets, thus mutations in miRNAs and their precursor molecules may affect the expression of multiple genes at once [142]. Examples of a variant within a miRNA precursor that affects its levels is the SNPs rs11614913 in miR-196a-2 (TT>CC), which leads to ineffective cleavage of the pre-miRNA in lung, breast, and gastric cancer (***Fig.10***) [143]. Additional examples are the rs7372209(CC>CT>TT) in 5' region of pri-miR62a-1 (putative targets include *E2F7, SMAD1* and *EZH2*) and rs1834306(A>G) in 5' region of pri-miR-100 (*HOXA* as a putative target), seemingly affecting their maturation processing and being correlated with increased drug resistance in metastatic colon cancer [144]. In some cancer types oncomiR loci can be duplicated due to CNVs, leading to increased expression levels. Such examples are the high expression levels of miR-296-5p, miR-3928-3p and miR-324-3p, which are correlated with poor prognosis in patients with Squamous cell lung carcinoma, as they target *FAM46C* leading to increased expression of *MYC* [145].

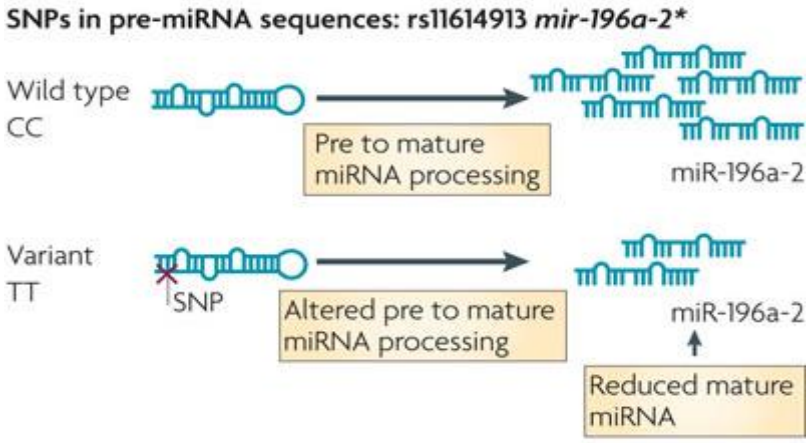

Figure 10: *SNP effect on pre-miRNA processing-case study of the rs11614913 SNP, which is part of the miR-196α-2 pre-miRNA. The wild-type CC allele is correlated with increased efficiency of the pre-miRNA processing, while the TT allele leads to reduced levels of the mature miRNA, correlated with increased risk for susceptibility for various forms of cancer.*

1.2.2.3.2 Role of lncRNAs and their mutations.

lncRNAs are transcripts that undergo capping and polyadenylation similarly to mRNAs [146]. They can be transcribed from a plethora of genomic areas and can act either *in cis* or *in trans* with respect to their target gene. They can be transcribed by intergenic or intronic areas, or by an alternative promoter site in parallel or anti-parallel of a protein-coding gene, while in some rare occasions circular lncRNAs (circRNAs) may occur by an exonic or intronic part of a gene [147]. One of the peculiarities of the lncRNAs' expression profile is that they show a high degree of specialization according to the cell-type and/or developmental/differentiation stage. They have many different functional roles within a cell, which can be further distinguished to their roles in the cytoplasm, and their roles within the nucleus [148, 149]. More specifically, in the nucleus:

1. They can interfere with chromatin writers that establish the epigenetic profile both in chromatin and DNA level, either by affecting positively or negatively their activity, or by targeting or blocking their localization to specific loci. These responses are specific for each lncRNA-protein interaction, as each lncRNA may alter the properties of its associating proteins in a different manner [150, 151].
2. They may interact with TFs and the RNA polymerase either by enhancing their affinity to interact with their binding sites, or by blocking their transcriptional role, as well as by stabilizing the promoter-enhancer loops, thus playing a key role in the regulation of the levels of transcription of a target gene [132, 152].
3. They promote intrachromosomal and interchromosomal interactions [153, 154].
4. They are responsible for the formation and action of the splicing machinery [155, 156].

On the other hand, cytoplasmic lncRNAs:

1. may act as miRNA sponges that have sequences that mimic those recognized by the miRNA in their mRNA targets. As a result, the miRNAs target the lncRNAs instead of the actual mRNA, thus enabling the latter to further be translated. In cancer, lncRNAs whose sites mimic those of oncogenes, are upregulated, thus the miRNAs target them instead of the oncogene, which in turn leads to increased protein levels of the latter. If their sites mimic those of a tumor-suppressor, then these lncRNAs are selectively down-regulated, leading to increased targeting of the tumor-suppressor in cancer cells [157, 158].
2. They interfere with the degradation of the mRNAs, either by interacting with the mRNA in a way that blocks the recognition sites of a miRNA, or by interacting with the proteins that are responsible for the degradation of the mRNA. Finally, in some cases lncRNAs are able to form hairpins, that can be identified and cleaved by Drosha, thus leading to the formation of a miRNA [159-161].
3. They may affect the translation of an mRNA, either by interacting with the ribosome, or with the mRNA, both in a negative or positive manner, depending on the lncRNAs' way of function [162].
4. They may produce functional small peptides, that are involved in a variety of biological processes [163].

Both in the cytoplasm and the nucleus, various cases of lncRNAs that enable interactions between different proteins have been found, especially in regard to molecular chaperones, mostly in a complex-stabilizing function (the roles of lncRNAs are summarized in ***Figure 11***) [164].

Two examples of SNPs that alter the function of lncRNAs in cancer are the SNPs rs6434568(C>A) and rs16834898(A>C) in PCGEM1, which have been associated with increased risk for prostate cancer appearance by enhancing the lncRNAs' ability to suppress the effect of drugs like doxorubicin, inhibiting cancer cell apoptosis [165]. An example of a SNP with a protective role, is SNP rs1317082(T>C) at exon 1 of CCSlnc362, which acts by creating a target site at the lncRNA for miR-4658, thus nullifying the lncRNA's ability to promote cellular proliferation and inhibition of apoptosis [166]. Investigating SNP-SNP interactions that lie within different lncRNAs can also be crucial, as it has been shown that interactions between 3 SNPs (rs17729428 TG+GG>TT, rs7958904 G>C and rs1899663 T>G) in the HOTAIR locus with the rs1859168(A>C) in the HOTTIP locus increase gastric cancer susceptibility through combinatorial miRNA target gain and loss [167]. Alterations in lncRNA expression levels have also been heavily linked with presence of CNVs, as 147 lncRNAs (such as *FENDRR*, which promotes the PRC2 mediated promoter methylation) have been shown to be downregulated in presence of CNVs in hepatocellular carcinoma [168].

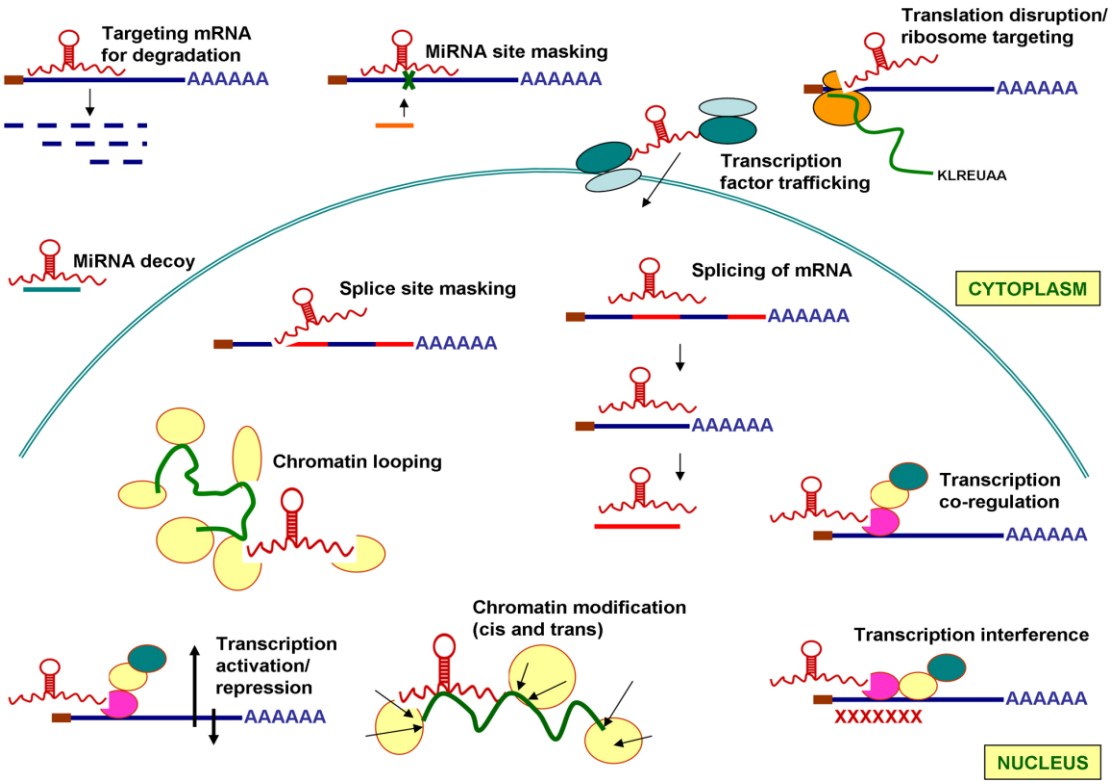

***Figure 11**: Roles of lncRNAs in the nucleus and the cytoplasm. In the nucleus lncRNAs are involved in chromatin epigenetic modifications, chromatin looping, transcription regulation and transcript splicing. In the cytoplasm, lncRNAs are involved in the translation procedure, and have multiple roles in miRNA-regulated mRNA degradation process. Both in the nucleus and the cytoplasm lncRNAs regulate protein interactions and trafficking of TFs.*

Thus, studying lncRNA regulatory function and expression profile together with their associated variants can be beneficial for finding new diagnostic tools and/or therapeutic targets with high specificity and sensitivity for various types of cancer, especially those that do not have any available biomarkers [169].

2. Aim of the study

This study focuses on the experimental validation of enhancer activity through the regulation neighboring lncRNA expression in gastric cancer. Apart from cancer specificity, these enhancer elements should be susceptible to cancer specific variation, due to the occurrence of somatic or germline SNPs. As a proof of concept during this study, we will create the necessary tools to gain insights on the regulatory effects of SNPs that reside within a putative enhancer element in the genomic locus of RECUR1, a lncRNA with putative oncogenic role in gastric cancer that was previously identified in our laboratory. This is achieved through the utilization of bioinformatic and experimental data, according to the following experimental pipeline :

1. First validate that the identified regulatory sequence act as an enhancer element of RECUR1 ex vivo,
2. Subsequently isolate the enhancer sequence from the genome of gastric cancer cells
3. Thirdly, create the necessary experimental tools for the functional characterization of the enhancer element along with its embedded genetic lesion , in order to study their regulatory capacity on RECUR1 transcription.

3. Materials and Methods

3.1 Cell culture and CRISPR activation system application

All sgRNAs were cloned in pLVU6 lentiviral vector (process shown in *Cloning* sector) and were used to create lentiviruses in modified HEK cells. The produced viruses were used to transfect gastric adenocarcinoma cell lines with a stable integration of dCas9 which was created in the lab (AGDC16).

3.2 RNA extraction and DNase treatment

Each RNA extraction was performed by using Trizol reagent (Tri reagent, MCA) with added chloroform for organic-aquatic phase separation and acquirement of the aquatic phase which contains the RNA part of the sample. Then, two subsequent isopropanol precipitation in presence of added glycogen and ethanol washing steps were performed to wash up the sample, and after removing the entire quantity of ethanol (pipetting & air drying), the RNA pellets were diluted in RNase free H_2O. All processes were performed in ice and the centrifugations in 4 ˚C (at maximum RPM), to avoid RNA degradation. Removal of residual DNA happens with a DNase reaction using DNaseI (ThermoFisher Scientific cat. #89836), according to the reaction protocol seen in ***Table1***.

Reagent	Quantity (30 µl reaction)	Duration 1h
10X DNaseI Buffer	3µl	Temperature 37˚C
DNaseI (1U/µl)	1µl	
RNase-out	0,5µl	
WFI	0,5 µl	
RNA sample	25µl (<10µg RNA)	

Table 1*: DNaseI reaction*

Following the DNaseI reaction, a phenol-chloroform-isoamyl alcohol phase separation is performed, and collection of the aquatic phase, in order to proceed with ethanol precipitation (100% ethanol) and an ethanol washing step (70% ethanol). After complete ethanol removal, dilution of the RNA pellet is performed with RNase-free water.

3.3 cDNA synthesis and qPCR

For the cDNA synthesis protocol, the reaction is performed following the instructions of the provider (Invitrogen Catalog number: 28025013) The cDNA synthesis reaction took place at 37 ˚C for 2 hours. For the qPCR reaction, the KAPA SYBRfast universal qPCR reaction buffer is used according to manufacturer instructions(Sigma Aldrich SKU KK4601). All qPCR reactions were run in Bio-Rad CFX Connect Real-Time PCR machine and analyzed using Bio-Rad CFX manager and MS Excel.

3.4 Cloning

In order to clone the enhancer sequence, a PCR standardization reaction using the Taq DNA polymerase from KAPA Biosystems was performed, with the following protocol:

Reagent	Enhancer (μL)	
Sample AGS gDNA	3,3	
10X Reaction Buffer	5	
dNTPs	1	
Primer RECUR1 Enh F1R1	2	
Taq DNA polymerase	0,3	
WFI	40,7	
Vfinal	50	
Step	**Duration**	**Temperature (°C)**
Initial Denaturation	3 min	95
Denaturation	30 sec	95
Annealing	20 sec	58
Extension	45 sec	72
Repeats	35	
Final extension	5 min	72
Hold	Infinite	12

***Table 2**: PCR protocol for parameter standardization.*

The product of the PCR reaction was run in an agarose gel (1% w/v agarose), in presence of EtBr for the imaging. The size of the expected amplicon is 696bp. Following standardization of the parameters required for obtaining the enhancer, a PCR using the Q5 High-Fidelity DNA polymerase is used, following the instructions provided by the company (New England Biolabs Catalog #M0491S -***Table 3***).

Reagent	Enhancer (μL)	
Sample AGS gDNA	3,3	
Buffer Q5 (5X)	10	
dNTPs	1	
Primer RECUR1 Enh F1R1	2	
Q5 polymerase	0,3	
WFI	33,4	
Vfinal	50	
Step	**Duration**	**Temperature (°C)**
Initial Denaturation	2 min	98
Denaturation	30 sec	98
Annealing	20 sec	58
Extension	45 sec	72
Repeats	39	
Final extension	2 min	72
Hold	Infinite	12

***Table 3**: PCR protocol using the Q5 DNA polymerase.*

Following the PCR reaction using the Q5 DNA polymerase, PCR clean-up was performed according to manufacture protocol (kit from Macherey-Nagel).

Subsequently the PCR product was subjected to a PNK reaction. In parallel, the cloning vector was cleaved using EcoRV (Enzyquest) for 5 hours at 37 °C, in order to create the blunt ends for the cloning. Following vector digestion, a clean-up step was used to

remove the restrictive endonuclease, followed by dephosphorylation of the digested vector using CIP (Calf Intestinal alkaline Phosphatase from New England Biolabs Catalog # M0290) prior to ligation. For the ligation, the T4 DNA ligase (New England Biolabs Catalog #M0202S) was used, with a vector to insert ratio of 1:3. All reactions were performed according to the instructions of the providing company.

The ligation product was used for transforming *E. coli* DH5α chemically competent cells. Single colonies were chosen for liquid cultivation (3ml, O/N cultures, 37˚C) and subsequent plasmid isolation by alkaline lysis to scan for positive colonies, alongside with insertion orientation. Proper insert cloning was verified by PCR using Taq DNA polymerase (***Table 4***), and the orientation of insertion by digestion using BsaI and HindIII-HF (New England Biolabs **Catalog** #R0535 & # R3104S -***Table 5***).

Reagents	**Quantity** (μL)	
Plasmid (10ng/μl)	1	
10X Buffer	3	
dNTPs	1	
Primer RECUR1 Enh F1R1	2	
Taq DNA polymerase	0,3	
WFI	22,7	
Vfinal	30	
Step	Duration	Temperature (°C)
Initial Denaturation	3 min	95
Denaturation	10 sec	95
Annealing	20 sec	58
Extension	45 sec	72
Repeats	30	
Final extension	5 min	72
Hold	Infinite	12

Table 4: *Diagnostic PCR protocol for insertion verification.*

Reagents	**Quantity (μl)**
Buffer	5
BsaI (20U/μl)	1,5
HindIII-HF (20U/μl)	1,5
Template	5
WFI	37
Vfinal	50

Table 5*: Diagnostic digestion protocol for insertion orientation using BsaI/HindIII.*

3.5 Genotyping of the SNP allele – dCAPS

To genotype the SNP in the cloned enhancer sequence, a dCAPS approach was selected alongside with SphI digestion. The conditions of the PCR reaction appear in ***Table 6***, while the digestion reaction with SphI (Enzyquest cat. #RE038S) appears in ***Table 7***.

Reagents	**Quantity (μL)**

Plasmid (10ng/μl)	1	
10X Buffer	5	
dNTPs	1	
Primer RECUR1 Enhancer dCAPS F1R1	2	
Taq DNA polymerase	0,3	
WFI	40,7	
Vfinal	50	
Step	Duration	Temperature (°C)
Initial Denaturation	5 min	95
Denaturation	30 sec	95
Annealing	20 sec	60
Extension	30 sec	72
Repeats	45	
Final extension	2 min	72
Hold	Infinite	12

***Table 6**: Diagnostic PCR protocol for dCAPS*

Reagents	Quantity
Buffer	5
SphI (10U/μl)	1,5
Template	20
WFI	23,5
Vfinal	50

***Table 7**: Digestion protocol for dCAPS.*

3.6 Site Directed Mutagenesis

Following SNP genotyping, the next step was to create an enhancer carrying the other allele (Mutant allele). To achieve this, a Site Directed Mutagenesis (SDM) protocol was performed. For this, 2 SDM PCR reactions (PCR 1 & PCR 2), each with an 40nt SDM primer with complementary sequence to each other were performed. The conditions of the reaction appear at ***Table 8*.**

Reagents	Quantity (μL)	
Plasmid (10ng/μl)	1	
5X Buffer	10	
dNTPs	1	
Primers (F1-SDM_R or SDM_F-R1)	1,5 for each primer	
Q5 DNA polymerase	0,25	
WFI	34,75	
Vfinal	50	
Step	Duration	Temperature (°C)
Initial Denaturation	3 min	98
Denaturation	10 sec	98
Annealing	10 sec	58 (PCR 1) / 60 (PCR 2)
Extension	20 sec	72
Repeats	20 (PCR 1) / 15 (PCR 2)	

Final extension	2 min	72
Hold	Infinite	12

***Table 8**: Conditions for the SDM PCRs 1 & 2 for creating the mutated allele.*

After successfully obtaining the first 2 PCRs of the protocol, the 2 PCR products can be used as a template, in order to recreate the complete enhancer sequence carrying the induced mutation via a third PCR reaction using the primers for the full-length cloning of the enhancer (***Table 9***)

Reagents	Quantity (μL)	
PCR_1 and PCR_2 products(diluted samples for equal molecular analogy)	1 each	
5X Buffer	10	
dNTPs	1	
Primers (Enhancer full length cloning F1R1)	1,5 for each primer	
Q5 DNA polymerase	0,25	
WFI	33,75	
Vfinal	50	
Step	Duration	Temperature (°C)
Initial Denaturation	3 min	98
Denaturation	10 sec	98
Annealing	10 sec	58
Extension	45 sec	72
Repeats	25	
Final extension	2 min	72
Hold	Infinite	12

***Table 9**: Conditions for the third PCR of the SDM reaction protocol for full length enhancer obtainment.*

SDM was verified again with dCAPS (described above) and the mutated enhancer sequence was ligated to the same vector in both orientations (*described above*). Final cloning products were verified base to base with Sanger sequencing performed with CeMIA.

4.Results

4.1 Bioinformatic analysis

The bioinformatic analysis (performed previously in the lab) showed two SNPs in the vicinity of the RECUR1 locus (REgulatory Cancer mUtation lncRNA 1). *RECUR1* transcribes a lncRNA with increased expression levels in patients with gastric cancer that was identified and initially characterized previously in our laboratory. The two SNPs reside i) upstream of the RECUR1 TSS and ii) within a nearby candidate enhancer element. Characterization of the regulatory role of the enhancer in RECUR1 expression alongside with the functional role of the embedded SNP may provide insights for gastric cancer predisposition alongside with applications in personalized diagnosis and therapy. This will be achieved through experimental validation via CRISPRa of the enhancer. Then, the enhancer sequence will be obtained by PCR and subsequently cloned to the appropriate vector. The presence of the allele will be verified by genotyping.

4.2 CRISPRa targeting the enhancer

Prior to downstream experiments, the regulatory role of the putative enhancer must be proven through correlation with increased RECUR1 expression. Verification of the regulatory capacity of the enhancer was performed with CRISPRa targeting the enhancer sequence. The RECUR1 promoter sequence was also included as a control. The first experiment of CRISPRa in AGdC16 cells (***Fig. 1)*** was not successful in overexpressing RECUR1, so a second experiment (***Fig. 2***) was performed using HdC1 (HEK epithelial cells expressing dCas9 clone 1), LdC10 (LS174 cells expressing dCas9 clone 10), which successfully overexpressed RECUR1.

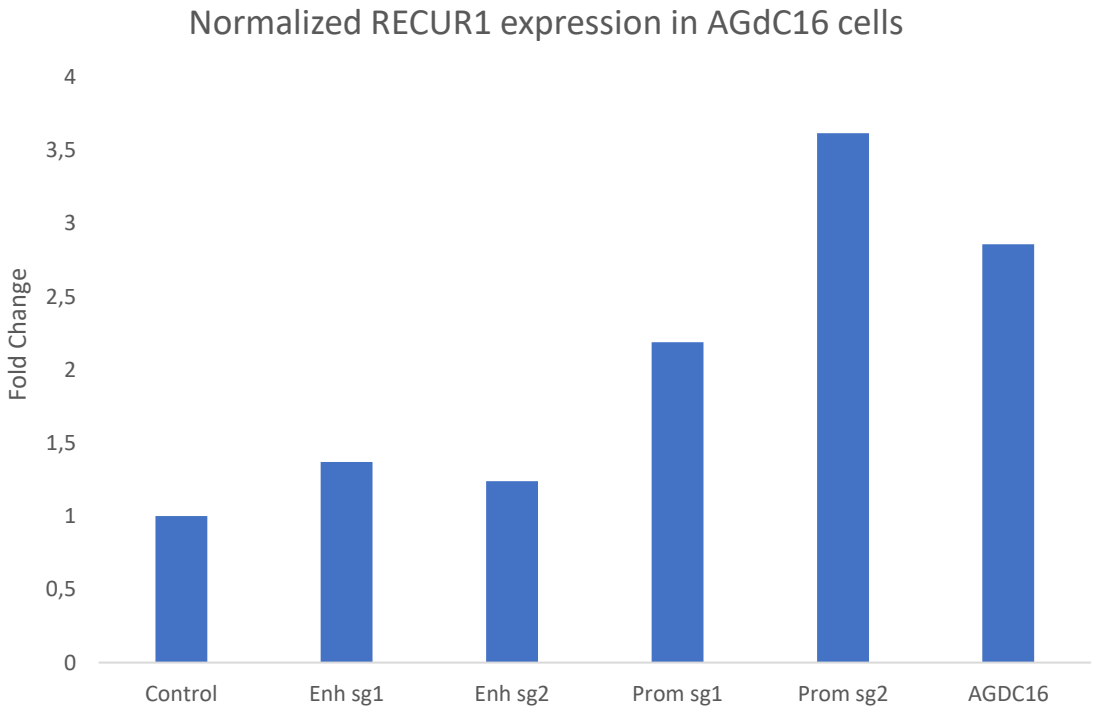

***Figure 1**: RECUR1 expression levels in AGdC16 cells. Overexpression of RECUR1 was not successful in the case of the sgRNAs targeting the enhancer in comparison to the expression levels achieved by targeting the promoter. All results were depicted as fold change in comparison to the control (empty vector).*

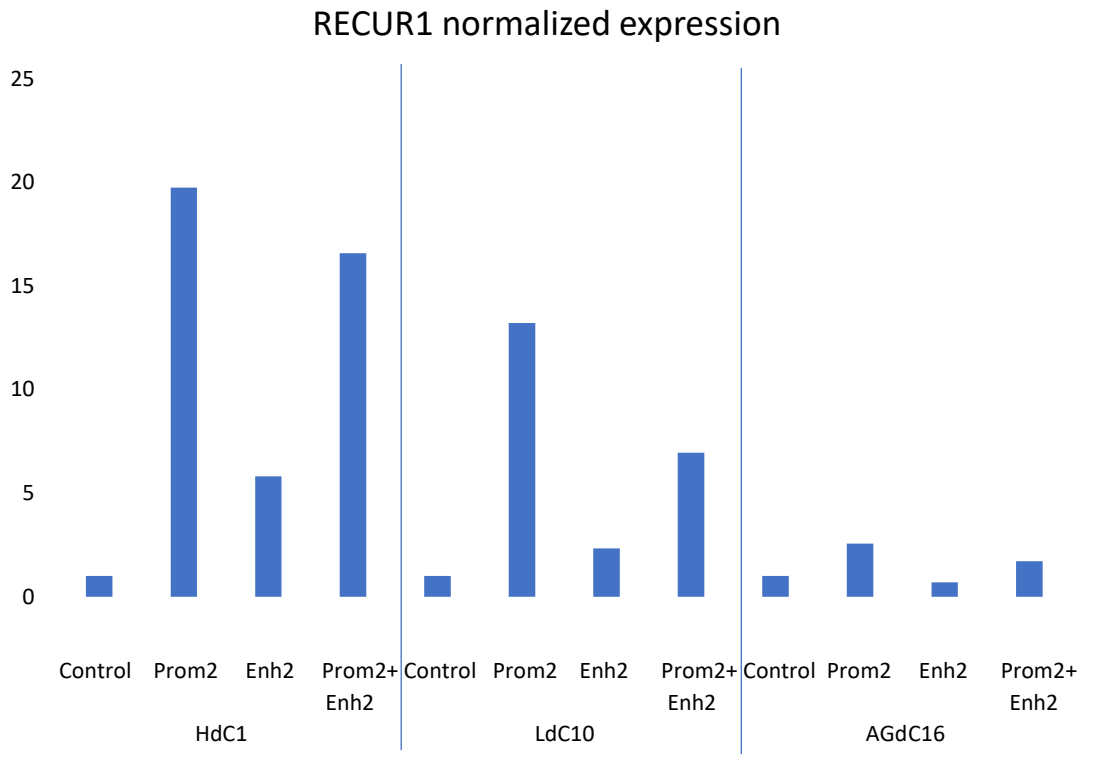

***Figure 2**: RECUR1 expression in HdC1, LdC10 and AGdC16 cells. Targeting of the enhancer using the sgRNA2 was successful in overexpressing RECUR1 in HdC1 and LdC10 cell lines, thus proving that the enhancer can regulate RECUR1. The sgRNA targeting the promoter showed increased expression of RECUR1 in all cell lines, while simultaneous targeting of the promoter and the enhancer managed to overexpress RECUR1, albeit in a smaller extent than single targeting of the promoter. All results for each cell line were depicted as fold change in comparison to the control (pLVU6).The diagram does not contain error bars, as it is based on one biological replicate due to time restrictions.*

As the two enhancer-targeting sgRNAs were unable to overexpress the lncRNA in a greater extent, additional sgRNAs were designed and successfully cloned to pLVU6[170] vector and are ready to be tested in the same cell lines (***Fig. 3***).

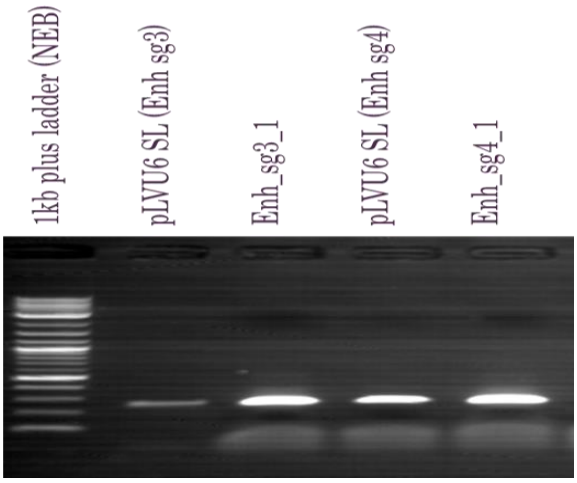

***Figure 3**: Diagnostic PCR indicating the successful insertion of the sgRNA3 and sgRNA4 with a sequence to target the enhancer regulating RECUR1. Electrophoresis performed on 1% Agarose gel.*

In the samples from the second experiment, another qPCR was performed the nearby protein coding genes to investigate alterations in their expression levels that could reflect regulation from the same enhancer element directly or indirectly by the lncRNA.

According to the results obtained by this experiment (***Fig.4***) a clear effect in the expression of both nearby genes was observed upon targeting of the RECUR1 promoter. On the other hand, the enhancer showed a steady profile in regulating the mRNA target 2 alongside with a tissue-specific potential for regulating mRNA target 1 depending on the cellular context.

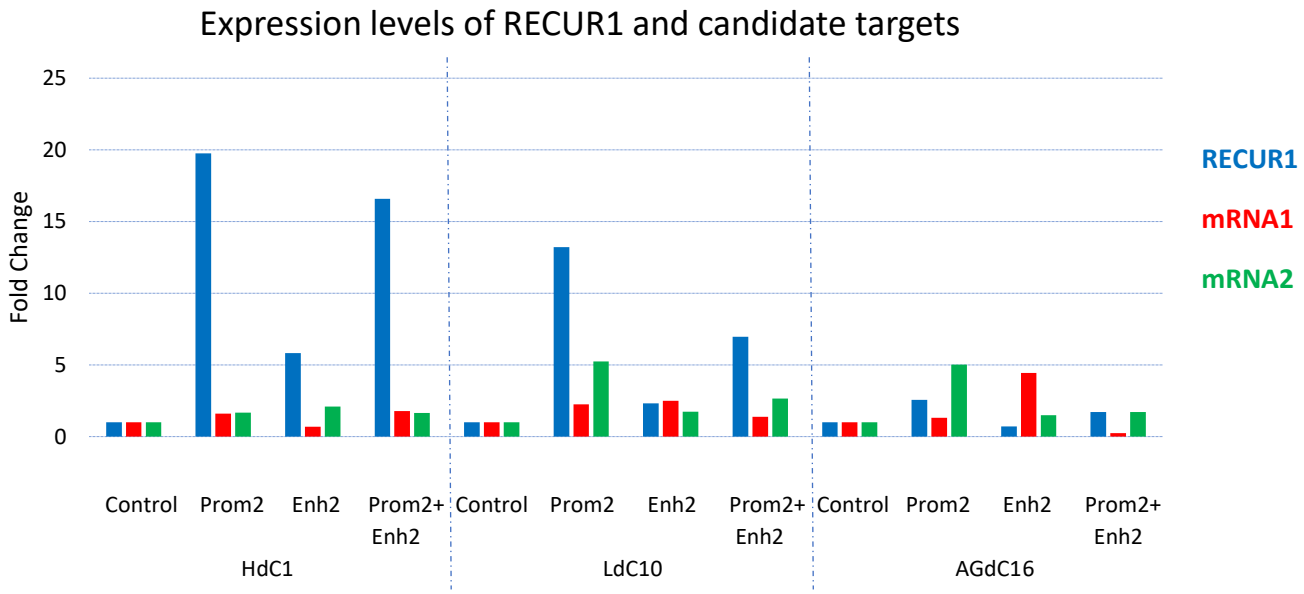

***Figure 4**: qPCR showing alterations in mRNA1 and mRNA2 levels. All results for each cell line were depicted as fold change in comparison to the control (pLVU6). In all samples, increased levels of mRNA2 were observed, for all target sites. mRNA1 expression alterations are target- and cell line-specific, as increase appears in all CRISPRa cases that are targeting the RECUR1 promoter. On the other hand, targeting of the enhancer results in decrease in HdC1 epithelial cells, yet increase in LdC10 and AGdC16 gastrointestinal cells. Simultaneous targeting of the promoter and the enhancer led to increased levels of mRNA1 in HdC1 and LdC10, but reduced expression levels in AGdC16.*

4.3 Cloning of the enhancer

After proving the regulatory potential of the enhancer element, the next step was to successfully obtain its sequence from genomic DNA (***Fig. 5***).

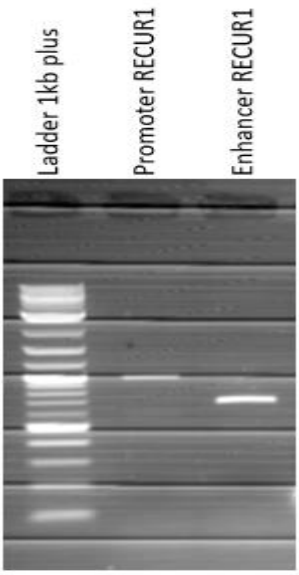

Figure 5*: Electrophoresis of the PCR used to obtain the full-length Promoter and Enhancer of RECUR1. Expected amplicon for the promoter → 1000bp, expected amplicon for the enhancer 696 bp. Electrophoresis performed on 1% Agarose gel.*

Following ligation to the vector and transformation to bacterial cells, single colonies were picked to extract the plasmids carrying the enhancer and promoter sequences (*PCR result in* ***Fig. 6***) in both insertion orientations (***Fig. 7****)*. The insertion orientation is required to verify whether each element acts indeed as a promoter (unidirectional function) and enhancer (bidirectional function).

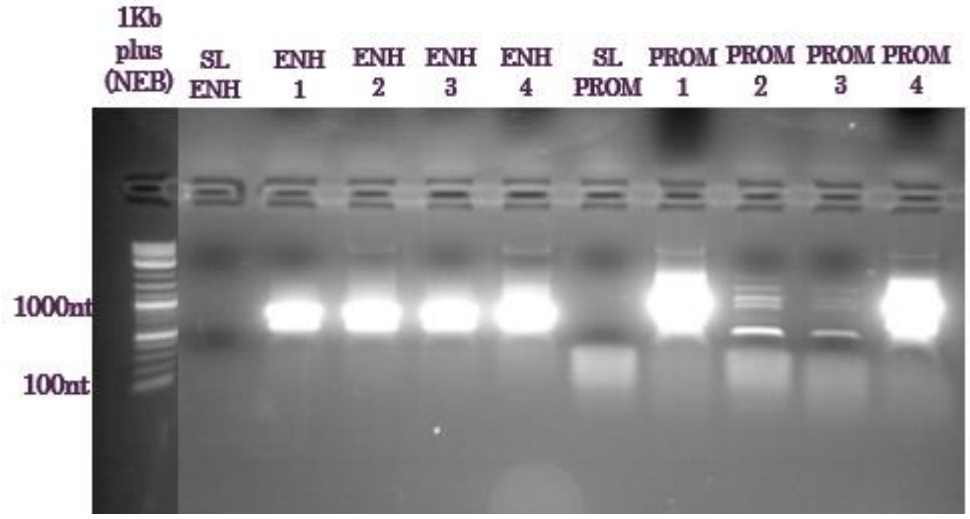

Figure 6*: Diagnostic PCR proving cloning of the promoter and the enhancer sequences. Expected amplicon for the promoter → 1000bp, expected amplicon for the enhancer 696 bp. Electrophoresis performed on 1% Agarose gel.*

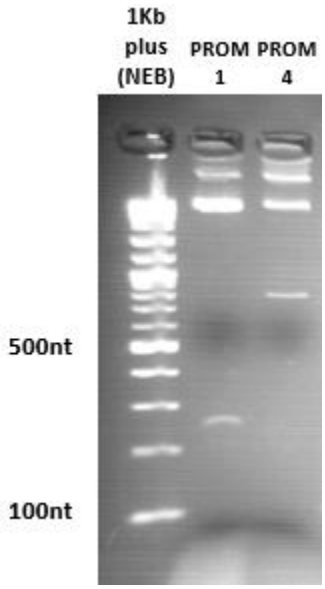

***Figure 7**: Digestion to clarify orientation of the promoter insertion. The expected products are: 5'-3'→ 230bp, 3'-5'→ 760bp. Electrophoresis performed on 1% Agarose gel.*

As seen in ***Figure 6***, the 1st and 4th clones of the promoter and all tested clones of the enhancer carried the insert. Luckily, the 1st clone of the promoter had the 5'-3' insertion orientation while the 4th clone of the promoter was inserted with the 3'-5' orientation (***Figure 7***). For the enhancer all initial positive clones had the same orientation, therefore additional colonies were picked and checked for the insertion orientation (***Fig. 8***).

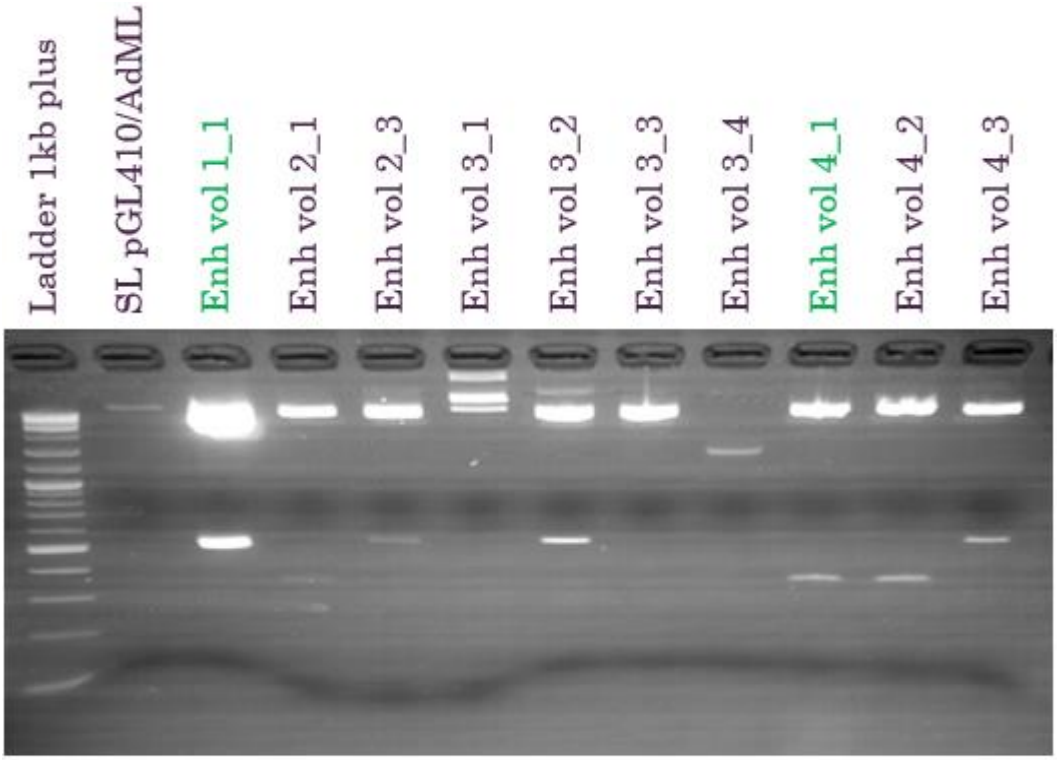

***Figure 8**: Digestion to clarify orientation of the enhancer. The expected products are: 5'-3'→ 497, 3'-5'→ 339. Green symbolizes the selected positive clones. Electrophoresis performed on 1% Agarose gel.*

The 1st clone from the first batch of clones picked (which contained the *5'-3'* insertion orientation), and the 1st clone from the 4th batch (which contained the *3'-5'* insertion orientation) were chosen for genotyping and subsequent Sanger sequencing (see below).

4.4 Genotyping and Site-Directed Mutagenesis

In order to validate the SNP allele in all cloned enhancer sequences, a dCAPS was performed. Digestion with SphI showed that both enhancer orientation clones selected in the previous step (see above, ***Fig 8***) carried the wild-type allele predicted to associate with reduced gastric cancer onset (***Fig. 9***).

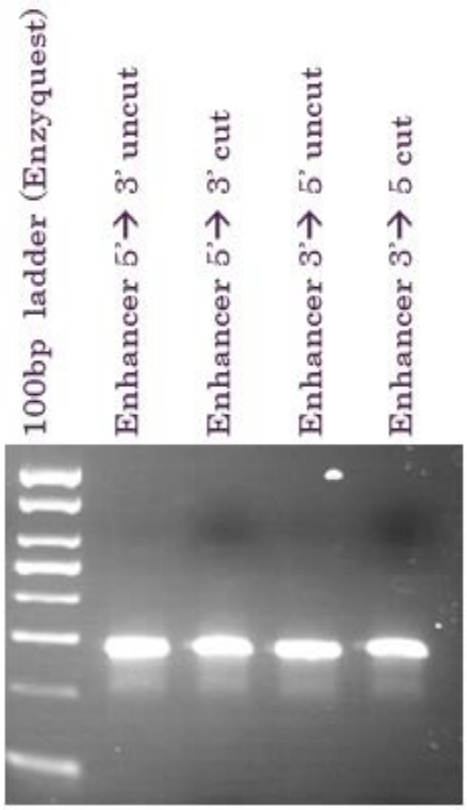

Figure 9*: Genotyping of the cloned enhancer sequence with the 5'-3'and 3'-5' insertion orientation. Uncut → the sample received after PCR and PCR clean-up and was not subjected to digestion. Cut → The digested PCR product. In this case, the enhancer carries the Wild-Type allele, which disrupts the recognition site for the restrictive endonuclease. PCR/Enhancer carrying the WT allele → 334 bp band. Electrophoresis performed on 2% Agarose gel.*

Following this result, a Site Directed Mutagenesis was performed for creating constructs harboring the alternative mutant allele according to [171]. PCR1 for the protocol was unsuccessful at the first attempt but was successfully obtained after some fine-tuning of the PCR conditions in the second attempt. PCR2 was successful on the first attempt (***Fig. 10***). The subsequent 3rd reaction for the SDM was duplicated to maximize the quantity of the final PCR product prior to ligation. The result of the successful attempt appears at ***Figure 11***.

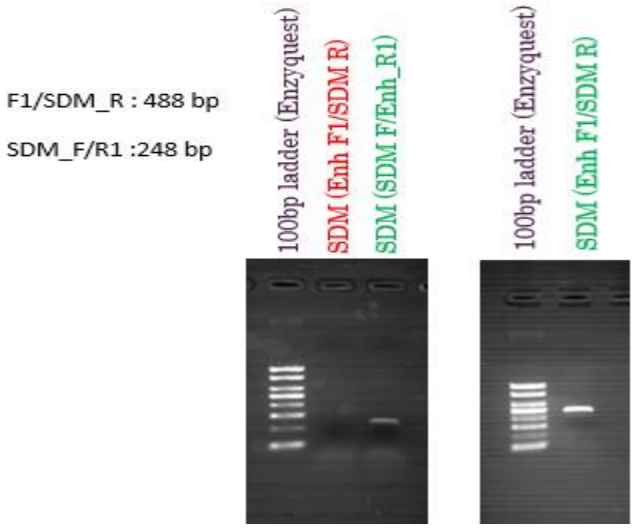

***Figure 10**: 1st (F1/SDM_R primers) and 2nd (SDM_F/R1 primers) PCR reactions for the SDM protocol. Red colorization → Unsuccessful first attempt at 1st PCR. Green colorization → Successful amplification of the PCR products for each PCR reaction. Electrophoresis performed on 1% Agarose gel.*

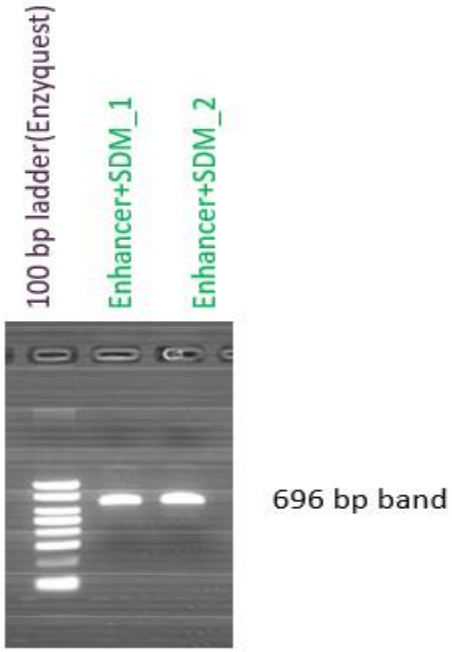

***Figure 11**: Successful recovery of the full-length enhancer fragment that carries the mutated allele. The expected band appears at 696 bp. The reaction was duplicated to have enough quantity prior to ligation. Electrophoresis performed on 1% Agarose gel.*

4.5 Cloning of the SDM products

A similar cloning approach was used to generate the constructs carrying the mutated allele as with the WT enhancer sequence. Following successful recovery of the mutated full-length enhancer (***Fig. 11***), 10 clones were picked to scan for positive insertion, as well as the insertion orientation, using a diagnostic double digestion with BsaI/HindIII (***Fig. 12***). After successfully obtaining the insert both in the *5'-3'* (2nd and 8th clones) and *3'-5'* (4th and 9th clones) insertion orientations, one of each orientation (2nd and 4th clone) were selected to proceed to genotyping and sequencing.

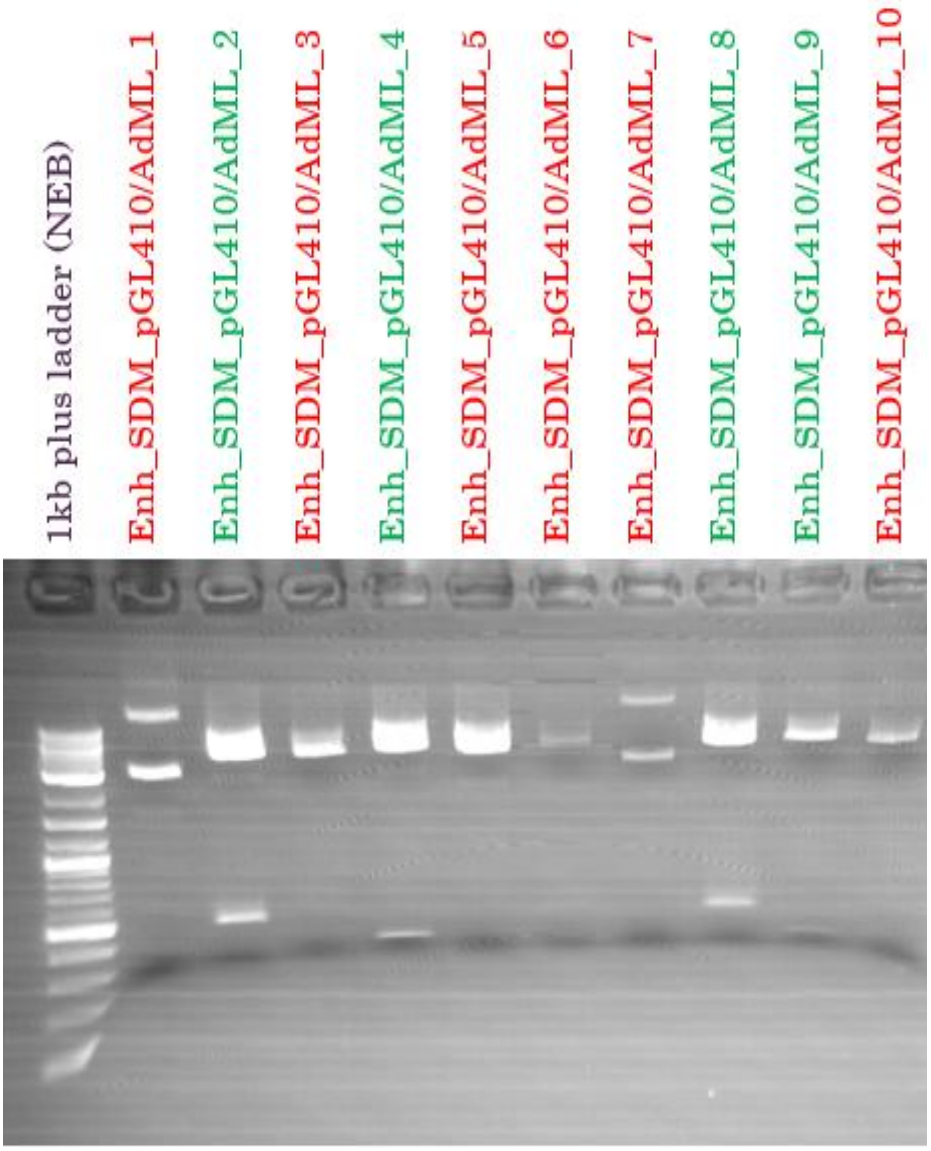

***Figure 12**: Diagnostic digestion using BsaI/HindIII to verify presence and insertion orientation of the insert. Red coloration → Negative clones for the insertion. Green coloration → Positive clones for the insertion. Expected bands: 5'-3' →497 bp, 3'-5'→339 bp. The 2nd and 8th samples carried the insert with 5'-3' insertion orientation and 4th and 9th clones carried the 3'-5' insertion orientation. The 2nd and 4th clones were selected for subsequent experiments. Electrophoresis performed on 1% Agarose gel.*

4.6 Genotyping of the clones carrying the SDM products

Using the selected enhancer clones from the previous step, the next goal was to verify whether the SDM conversion was successful, and the cloned enhancer sequence indeed carried the mutated allele for the SNP of focus. To do that, a dCAPS approach [172] was performed (same strategy as in section *4.4*) using SphI, which only cleaves in presence of the mutated allele. Indeed, the SDM mutagenesis was successful, as the enzyme was able to cleave (partial digestion due to excess amount of PCR product in the reaction) the enhancer sequence at the site of the allele, generating a 294bp and a 40 bp band (***Fig. 13***).

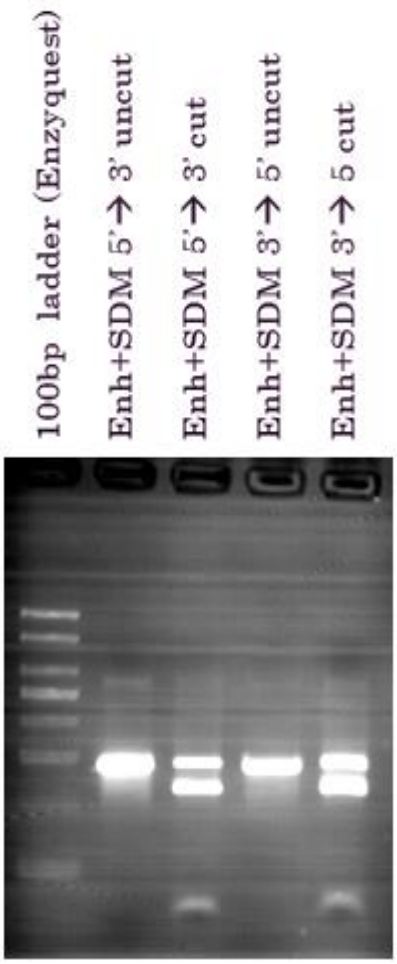

***Figure 13**: Genotyping of the enhancer sequence carried in the two plasmids with the 5'-3'and 3'-5' insertion orientation. The expected bands are for the PCR/Enhancer carrying the WT allele → 334 bp band, and for the digested enhancer with the mutated allele → 294bp and 40bp band. Uncut → the sample received after PCR and PCR clean-up and was not subjected to digestion. Cut → The digested PCR product. The result shows that partial digestion occurred in the restriction digest reaction with SphI, indicating: 1) excess amount of PCR product present in the sample, that did not allow complete digestion and 2) the presence of the mutated allele which creates the cleaving site for SphI, thus concluding that the SDM conversion was successful. Electrophoresis performed on 2% Agarose gel.*

These samples that carried the mutated allele were sent for sequencing, along with the samples carrying the WT allele. Sequencing results proved that each individual clone from both experiments (clones obtained with both orientations with and without SDM conversion) indeed carried the respective alleles, thus verifying the dCAPS results (***Fig. 14***). Sequencing results also verified the insertion and orientation, as shown with restriction digest, as well as a 100% sequence match to the enhancer sequence compared to the human genome assembly (with the exception of the mutated SNP allele). Thus, those clones could now be used for a luciferase assay to test the regulatory capacity of the enhancer and the SNP contained within it.

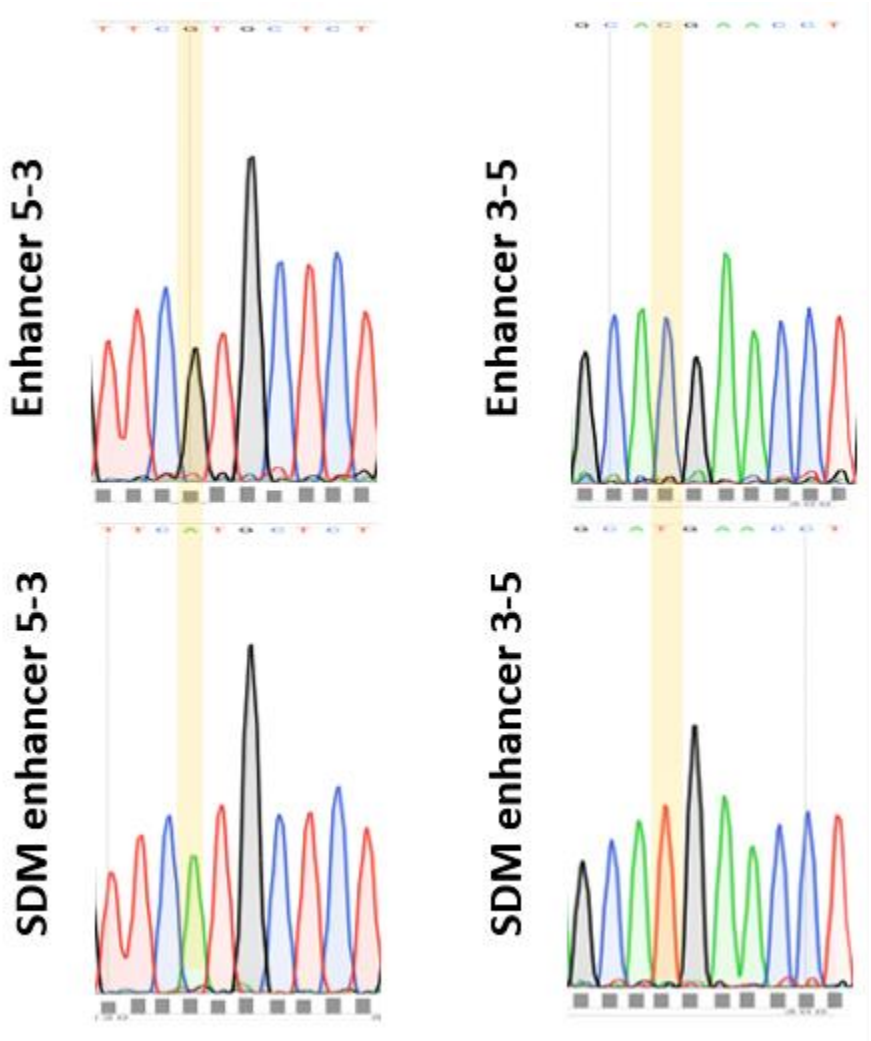

***Figure 14**: Sequencing results, focused on the adjacent area of the SNP within the enhancer. The SNP allele in each sample appears highlighted with yellow colorization. All samples carried the expected allele based on the previous results.*

5.Discussion

SNPs are major players in disease onset [173]. Their presence has been correlated with deregulated protein function, alterations in ncRNA mechanism and can drastically affect the function of DNA regulatory elements [174-176]. In the presence of coding SNPS, protein function can be affected either through altered amino-acid sequence or through alternative splicing (intron retention, exon exclusion) [177]. On a regulatory level, altered protein levels may occur due to SNP presence within the transcripts' UTRs, the miRNAs that target them, or by alterations in the proteins' non-transcribed regulatory machinery [178]. SNPs might alter the binding affinity of transcription factors at regulatory elements [179]. Additionally, studies have shown that presence of SNPs affect the function of ncRNAs, either by affecting their maturation process or their secondary structure[180].

Focusing on the effect of SNPs in enhancer regions, many studies have shown their effect in deregulating both coding and non-coding RNA expression. Such an example is the SNP rs1892901, a disease-related SNP detected in high occurrence in Chinese populations. This SNP is correlated with multiple forms of cancers, most prominently, gastric. It is situated near an enhancer element regulating *FOSL1* (FOS-like antigen 1), a proto-oncogene with pivotal role in cell proliferation, differentiation, and metastasis. This occurs as FOSL1 has been identified as a deregulator of PI3K/Akt and p53 pathways in gastric cancer cells [181]. Analyses showed that rs1892901 can be part of transcription factor binding sites, and affect the binding affinity of multiple transcription factors, such as EGR1, CHD, EP300, FOS, JUN, FOSL1, and FOSL2 [182]. All these transcription factors have a well-established role in gastric cancer appearance [183-186].

Two SNPs with a high risk for colorectal cancer are rs10505477 and rs10411210 [187]. rs10505477, in particular, has also been corelated with increased risk for intestinal type of gastric cancer, affecting the patients' survival and treatment response [188]. These SNPs lie within super-enhancers bound by CDX2 and HNF4α, two oncogenic transcription factors known to regulate colonic development [189, 190]. The risk alleles of both SNPs are correlated with increased levels of $H_3K_{27}ac$ at their respective super-enhancers in comparison to normal alleles, thus proving a correlation between histone modifications and DNA sequence [191].

HOTAIR is one of the most well-characterized lncRNAs, and its role is highlighted in gastric cancer, as well as other cancer forms, acting as a predictor of the patients' survival [192-194]. Mechanistically, it interacts with Polycomb Repressive Complex 2 (PRC2) and lysine-specific demethylase 1 (LSD1), targeting them to certain genomic loci. As a result, multiple tumor-suppressor genes, like *JAM2*, *PCDH10* and *PCDHB5* are downregulated, while several oncogenes like SNAIL are overexpressed [195]. Multiple SNPs have been found regulating the expression of HOTAIR in gastric cancer (along with many other forms). rs920778 lies within an intronic enhancer of *HOTAIR*, has been a heavily disputed SNP correlated with increased expression of HOTAIR in gastric cancer and esophageal squamous cell carcinoma [196, 197]. On the other hand, the

high prostate and lung cancer risk SNP rs1899663 is located in a putative regulatory element within *HOTAIR*, and alters the binding affinity of PAX-4, SPZ1, and ZFP281 through altering the chromatin accessibility in the region [198].

Another lncRNA regulated by a SNP lying within an enhancer element, it PCAT1 (Prostate Cancer Associated Transcript 1). PCAT1 is directly involved in prostate cancer cell proliferation and castration resistance, while it has been correlated with gastric cancer appearance. The high-risk T allele of the SNP rs7463708 (T>G) increases the binding affinity of ONECUT transcription factor family members (like HNF4 and HNF6) and AR (androgen receptor). In turn, the PCAT1 transcript interacts with LSD1 and AR, targeting them to enhancers regulating *GNMT* and *DHCR24,* two genes with established role in prostate cancer progression [199-203]. In gastric cancer, in particular, PCAT1 upregulates the expression of CDKN1A, leading to increased cell proliferation, migration, and invasion [204].

A novel enhancer studied during this thesis, is an enhancer regulating RECUR1. The *in-silico* analysis showed the SNP lying within this enhancer sequence has a positive correlation with gastric cancer appearance. Its role in enhancer function is yet to be found, as well as the way it may affect the binding of KLF5 to the enhancer element. The next step of the analysis is to verify how the mutated dominant allele may affect the functionality of the enhancer in an episomal assay *in cellulo*. But, since this is an episomal assay that studies the functionality of the enhancer away of its natural habitat within its specific genomic locus, another assay needs to be performed, which would alter the SNP allele on genomic DNA level. Such an approach is a CRISPR editing with a homologous recombination approach targeted at the site of the SNP. This would require the conduction of an experiment monitoring RECUR1 expression both in cells homozygous for the recessive allele, as well as cells homozygous for the dominant-cancer related allele. This experiment would allow the characterization of the SNP as a driver or passenger variant. Depending on the result (both from the episomal and the genome editing assays), multiple experiments can be performed to test transcription factor binding to the enhancer and local chromatin architecture. A chromatin immunoprecipitation using antibodies for KLF5 followed by qPCR (ChIP-qPCR protocol) can be used with chromatin from cells carrying each allele in a homozygous state. This experiment would provide information about the binding ability of KLF5 in presence of each allele. Additional experiments may also focus on detection of other transcription factors that may also bind in this enhancer, but this would require extra analysis of metagenomic data coupled with immunoprecipitation and Mass Spectrometry for their detection. Another experiment that can be performed is a chromosome confirmation capture (3C), possibly coupled with qPCR [205]. This protocol would correlate the genome architecture through loop formation in the genome locus with the presence of each allele. This would lead to a complete mechanistical characterization of the SNP and its effect on enhancer activity.

Ultimately, if this SNP indeed proves to have an allele-specific correlation with increased risk for gastric cancer appearance, it would be vastly beneficial for a

prognostic and diagnostic tool development. More specifically, such tools would provide an easy and cost-efficient method to detect an individual's probability to succumb to the disease [206]. This allows a better monitoring of each person within a population even at a young age, prior to appearance of the disease to receive precaution measures, or the disease will be detected at an early stage, in which it is still curable. Depending on the results from further analyses focused on the enhancer function, along with the role of the lncRNA, it may enable their therapeutic targeting. This might be achieved either through pharmaceutical targeting focused at the lncRNA, or through genome editing focusing on the enhancer or the lncRNA[207].

6. Citations

1. Ajani, J.A., et al., *Gastric adenocarcinoma.* Nat Rev Dis Primers, 2017. **3**: p. 17036.
2. Choi, K.S., et al., *Effect of endoscopy screening on stage at gastric cancer diagnosis: results of the National Cancer Screening Programme in Korea.* Br J Cancer, 2015. **112**(3): p. 608-12.
3. Khanderia, E., et al., *The Influence of Gastric Cancer Screening on the Stage at Diagnosis and Survival: A Meta-Analysis of Comparative Studies in the Far East.* J Clin Gastroenterol, 2016. **50**(3): p. 190-7.
4. Chen, W.Y., et al., *Factors that affect life expectancy of patients with gastric adenocarcinoma.* Clin Gastroenterol Hepatol, 2013. **11**(12): p. 1595-600.
5. Chu, P.C., et al., *Estimation of life expectancy and the expected years of life lost in patients with major cancers: extrapolation of survival curves under high-censored rates.* Value Health, 2008. **11**(7): p. 1102-9.
6. Apicella, M., S. Corso, and S. Giordano, *Targeted therapies for gastric cancer: failures and hopes from clinical trials.* Oncotarget, 2017. **8**(34): p. 57654-57669.
7. Pellino, A., et al., *Targeted therapies in metastatic gastric cancer: Current knowledge and future perspectives.* World J Gastroenterol, 2019. **25**(38): p. 5773-5788.
8. Richman, D.M., et al., *Beyond gastric adenocarcinoma: Multimodality assessment of common and uncommon gastric neoplasms.* Abdom Radiol (NY), 2017. **42**(1): p. 124-140.
9. Padmanabhan, N., T. Ushijima, and P. Tan, *How to stomach an epigenetic insult: the gastric cancer epigenome.* Nat Rev Gastroenterol Hepatol, 2017. **14**(8): p. 467-478.
10. Lazar, D.C., et al., *Prognostic significance of tumor immune microenvironment and immunotherapy: Novel insights and future perspectives in gastric cancer.* World J Gastroenterol, 2018. **24**(32): p. 3583-3616.
11. Wang, Q., G. Liu, and C. Hu, *Molecular Classification of Gastric Adenocarcinoma.* Gastroenterology Res, 2019. **12**(6): p. 275-282.
12. Zhang, W., *TCGA divides gastric cancer into four molecular subtypes: implications for individualized therapeutics.* Chin J Cancer, 2014. **33**(10): p. 469-70.
13. Cancer Genome Atlas Research, N., *Comprehensive molecular characterization of gastric adenocarcinoma.* Nature, 2014. **513**(7517): p. 202-9.
14. Oh, S.C., et al., *Clinical and genomic landscape of gastric cancer with a mesenchymal phenotype.* Nat Commun, 2018. **9**(1): p. 1777.
15. Cao, Z., et al., *Association between polymorphisms in XRCC1 gene and treatment outcomes of patients with advanced gastric cancer: a systematic review and meta-analysis.* PLoS One, 2014. **9**(1): p. e85357.
16. Gattas, G.J., et al., *Genetic polymorphisms of CYP1A1, CYP2E1, GSTM1, and GSTT1 associated with head and neck cancer.* Head Neck, 2006. **28**(9): p. 819-26.
17. Lin, Y., et al., *Correlation between SNPs in CDH1 and gastric cancer in Chinese population.* Open Med (Wars), 2015. **10**(1): p. 57-62.
18. Moradi, M.T., et al., *Effects of p53 codon 72 and MDM2 SNP309 polymorphisms on gastric cancer risk among the Iranian population.* Asian Pac J Cancer Prev, 2014. **15**(17): p. 7413-7.
19. Yuan, J., et al., *Risk prediction for early-onset gastric carcinoma: a case-control study of polygenic gastric cancer in Han Chinese with hereditary background.* Oncotarget, 2016. **7**(23): p. 33608-15.
20. He, C., et al., *SNP interactions of Helicobacter pylori-related host genes PGC, PTPN11, IL1B, and TLR4 in susceptibility to gastric carcinogenesis.* Oncotarget, 2015. **6**(22): p. 19017-26.

21. Sung, H., et al., *Association of high-evidence gastric cancer susceptibility loci and somatic gene expression levels with survival.* Carcinogenesis, 2017. **38**(11): p. 1119-1128.
22. Gu, Y., et al., *PSCA s2294008 C>T and rs2976392 G>A polymorphisms contribute to cancer susceptibility: evidence from published studies.* Genes Cancer, 2015. **6**(5-6): p. 254-264.
23. Ni, J., et al., *Correlation between protein kinase catalytic subunit alpha-1 gene rs13361707 polymorphism and gastric cancer susceptibility in asian populations.* Oncotarget, 2017. **8**(40): p. 68354-68364.
24. Maleki, S.S. and C. Rocken, *Chromosomal Instability in Gastric Cancer Biology.* Neoplasia, 2017. **19**(5): p. 412-420.
25. McLean, M.H. and E.M. El-Omar, *Genetics of gastric cancer.* Nat Rev Gastroenterol Hepatol, 2014. **11**(11): p. 664-74.
26. Sohn, B.H., et al., *Clinical Significance of Four Molecular Subtypes of Gastric Cancer Identified by The Cancer Genome Atlas Project.* Clin Cancer Res, 2017.
27. Abrahao-Machado, L.F. and C. Scapulatempo-Neto, *HER2 testing in gastric cancer: An update.* World J Gastroenterol, 2016. **22**(19): p. 4619-25.
28. Boku, N., *HER2-positive gastric cancer.* Gastric Cancer, 2014. **17**(1): p. 1-12.
29. Tao, J., et al., *CD44-SLC1A2 gene fusions in gastric cancer.* Sci Transl Med, 2011. **3**(77): p. 77ra30.
30. Polom, K., et al., *Meta-analysis of microsatellite instability in relation to clinicopathological characteristics and overall survival in gastric cancer.* Br J Surg, 2018. **105**(3): p. 159-167.
31. Ratti, M., et al., *Microsatellite instability in gastric cancer: molecular bases, clinical perspectives, and new treatment approaches.* Cell Mol Life Sci, 2018. **75**(22): p. 4151-4162.
32. Muratani, M., et al., *Nanoscale chromatin profiling of gastric adenocarcinoma reveals cancer-associated cryptic promoters and somatically acquired regulatory elements.* Nat Commun, 2014. **5**: p. 4361.
33. Guo, Y.A., et al., *Mutation hotspots at CTCF binding sites coupled to chromosomal instability in gastrointestinal cancers.* Nat Commun, 2018. **9**(1): p. 1520.
34. Hnisz, D., et al., *Super-enhancers in the control of cell identity and disease.* Cell, 2013. **155**(4): p. 934-47.
35. He, F., et al., *Integrative Analysis of Somatic Mutations in Non-coding Regions Altering RNA Secondary Structures in Cancer Genomes.* Sci Rep, 2019. **9**(1): p. 8205.
36. Uhler, C. and G.V. Shivashankar, *Regulation of genome organization and gene expression by nuclear mechanotransduction.* Nat Rev Mol Cell Biol, 2017. **18**(12): p. 717-727.
37. Berman, B.P., et al., *Regions of focal DNA hypermethylation and long-range hypomethylation in colorectal cancer coincide with nuclear lamina-associated domains.* Nat Genet, 2011. **44**(1): p. 40-6.
38. Gopi, L.K. and B.L. Kidder, *Integrative pan cancer analysis reveals epigenomic variation in cancer type and cell specific chromatin domains.* Nat Commun, 2021. **12**(1): p. 1419.
39. Ezkurdia, I., et al., *Multiple evidence strands suggest that there may be as few as 19,000 human protein-coding genes.* Hum Mol Genet, 2014. **23**(22): p. 5866-78.
40. Salzberg, S.L., *Open questions: How many genes do we have?* BMC Biol, 2018. **16**(1): p. 94.
41. Kvon, E.Z., et al., *Enhancer redundancy in development and disease.* Nat Rev Genet, 2021.
42. Liu, M., et al., *Genomic discovery of potent chromatin insulators for human gene therapy.* Nat Biotechnol, 2015. **33**(2): p. 198-203.

43. Vitsios, D., et al., *Prioritizing non-coding regions based on human genomic constraint and sequence context with deep learning.* Nat Commun, 2021. **12**(1): p. 1504.
44. Kulaeva, O.I., et al., *Distant activation of transcription: mechanisms of enhancer action.* Mol Cell Biol, 2012. **32**(24): p. 4892-7.
45. Pindyurin, A.V., J. de Jong, and W. Akhtar, *TRIP through the chromatin: a high throughput exploration of enhancer regulatory landscapes.* Genomics, 2015. **106**(3): p. 171-177.
46. Genomes Project, C., et al., *A global reference for human genetic variation.* Nature, 2015. **526**(7571): p. 68-74.
47. Genomes Project, C., et al., *An integrated map of genetic variation from 1,092 human genomes.* Nature, 2012. **491**(7422): p. 56-65.
48. International HapMap, C., *The International HapMap Project.* Nature, 2003. **426**(6968): p. 789-96.
49. International HapMap, C., et al., *A second generation human haplotype map of over 3.1 million SNPs.* Nature, 2007. **449**(7164): p. 851-61.
50. Fagny, M., et al., *Nongenic cancer-risk SNPs affect oncogenes, tumour-suppressor genes, and immune function.* Br J Cancer, 2020. **122**(4): p. 569-577.
51. Rashkin, S.R., et al., *Pan-cancer study detects genetic risk variants and shared genetic basis in two large cohorts.* Nat Commun, 2020. **11**(1): p. 4423.
52. Sieh, W., et al., *Identification of 31 loci for mammographic density phenotypes and their associations with breast cancer risk.* Nat Commun, 2020. **11**(1): p. 5116.
53. Yao, L., et al., *Functional annotation of colon cancer risk SNPs.* Nat Commun, 2014. **5**: p. 5114.
54. Peyrot, W.J. and A.L. Price, *Identifying loci with different allele frequencies among cases of eight psychiatric disorders using CC-GWAS.* Nat Genet, 2021.
55. Wu, Y., et al., *GWAS of peptic ulcer disease implicates Helicobacter pylori infection, other gastrointestinal disorders and depression.* Nat Commun, 2021. **12**(1): p. 1146.
56. Slatkin, M., *Linkage disequilibrium--understanding the evolutionary past and mapping the medical future.* Nat Rev Genet, 2008. **9**(6): p. 477-85.
57. Carrasco Pro, S., et al., *Prediction of genome-wide effects of single nucleotide variants on transcription factor binding.* Sci Rep, 2020. **10**(1): p. 17632.
58. Lin, Y., et al., *Genomic and transcriptomic alterations associated with drug vulnerabilities and prognosis in adenocarcinoma at the gastroesophageal junction.* Nat Commun, 2020. **11**(1): p. 6091.
59. Liu, L.Y., et al., *Quantifying the influence of mutation detection on tumour subclonal reconstruction.* Nat Commun, 2020. **11**(1): p. 6247.
60. Kumaran, M., et al., *Breast cancer associated germline structural variants harboring small noncoding RNAs impact post-transcriptional gene regulation.* Sci Rep, 2018. **8**(1): p. 7529.
61. Li, Y.R., et al., *Rare copy number variants in over 100,000 European ancestry subjects reveal multiple disease associations.* Nat Commun, 2020. **11**(1): p. 255.
62. Shi, X., et al., *Association of CNVs with methylation variation.* NPJ Genom Med, 2020. **5**: p. 41.
63. Conrad, D.F., et al., *Origins and functional impact of copy number variation in the human genome.* Nature, 2010. **464**(7289): p. 704-12.
64. Holt, R., et al., *CNVs leading to fusion transcripts in individuals with autism spectrum disorder.* Eur J Hum Genet, 2012. **20**(11): p. 1141-7.
65. Mamlouk, S., et al., *DNA copy number changes define spatial patterns of heterogeneity in colorectal cancer.* Nat Commun, 2017. **8**: p. 14093.
66. Truty, R., et al., *Prevalence and properties of intragenic copy-number variation in Mendelian disease genes.* Genet Med, 2019. **21**(1): p. 114-123.

67. Zhang, D., et al., *Intron retention is a hallmark and spliceosome represents a therapeutic vulnerability in aggressive prostate cancer.* Nat Commun, 2020. **11**(1): p. 2089.
68. Group, P.T.C., et al., *Genomic basis for RNA alterations in cancer.* Nature, 2020. **578**(7793): p. 129-136.
69. Zhu, H., et al., *Candidate Cancer Driver Mutations in Distal Regulatory Elements and Long-Range Chromatin Interaction Networks.* Mol Cell, 2020. **77**(6): p. 1307-1321 e10.
70. Morlando, M., M. Ballarino, and A. Fatica, *Long Non-Coding RNAs: New Players in Hematopoiesis and Leukemia.* Front Med (Lausanne), 2015. **2**: p. 23.
71. Jeon, Y.J., et al., *Genetic variants in 3'-UTRs of methylenetetrahydrofolate reductase (MTHFR) predict colorectal cancer susceptibility in Koreans.* Sci Rep, 2015. **5**: p. 11006.
72. Bicknell, A.A., et al., *Introns in UTRs: why we should stop ignoring them.* Bioessays, 2012. **34**(12): p. 1025-34.
73. Cenik, C., et al., *Genome analysis reveals interplay between 5'UTR introns and nuclear mRNA export for secretory and mitochondrial genes.* PLoS Genet, 2011. **7**(4): p. e1001366.
74. Rearick, D., et al., *Critical association of ncRNA with introns.* Nucleic Acids Res, 2011. **39**(6): p. 2357-66.
75. Rose, A.B., *Intron-mediated regulation of gene expression.* Curr Top Microbiol Immunol, 2008. **326**: p. 277-90.
76. Bonnet, A., et al., *Introns Protect Eukaryotic Genomes from Transcription-Associated Genetic Instability.* Mol Cell, 2017. **67**(4): p. 608-621 e6.
77. Kahles, A., et al., *Comprehensive Analysis of Alternative Splicing Across Tumors from 8,705 Patients.* Cancer Cell, 2018. **34**(2): p. 211-224 e6.
78. Jung, H., K.S. Lee, and J.K. Choi, *Comprehensive characterisation of intronic mis-splicing mutations in human cancers.* Oncogene, 2021. **40**(7): p. 1347-1361.
79. Oltean, S. and D.O. Bates, *Hallmarks of alternative splicing in cancer.* Oncogene, 2014. **33**(46): p. 5311-8.
80. Bailey, M.H., et al., *Retrospective evaluation of whole exome and genome mutation calls in 746 cancer samples.* Nat Commun, 2020. **11**(1): p. 4748.
81. Liu, Y., et al., *Rare deleterious germline variants and risk of lung cancer.* NPJ Precis Oncol, 2021. **5**(1): p. 12.
82. Arthur, S.E., et al., *Genome-wide discovery of somatic regulatory variants in diffuse large B-cell lymphoma.* Nat Commun, 2018. **9**(1): p. 4001.
83. Diederichs, S., et al., *The dark matter of the cancer genome: aberrations in regulatory elements, untranslated regions, splice sites, non-coding RNA and synonymous mutations.* EMBO Mol Med, 2016. **8**(5): p. 442-57.
84. Dai, J., et al., *Systematical analyses of variants in CTCF-binding sites identified a novel lung cancer susceptibility locus among Chinese population.* Sci Rep, 2015. **5**: p. 7833.
85. J, V., et al., *Genetic association of MMP14 promoter variants and their functional significance in gallbladder cancer pathogenesis.* J Hum Genet, 2021.
86. Schodel, J., et al., *Common genetic variants at the 11q13.3 renal cancer susceptibility locus influence binding of HIF to an enhancer of cyclin D1 expression.* Nat Genet, 2012. **44**(4): p. 420-5, S1-2.
87. Haberle, V. and A. Stark, *Eukaryotic core promoters and the functional basis of transcription initiation.* Nat Rev Mol Cell Biol, 2018. **19**(10): p. 621-637.
88. Noe Gonzalez, M., D. Blears, and J.Q. Svejstrup, *Causes and consequences of RNA polymerase II stalling during transcript elongation.* Nat Rev Mol Cell Biol, 2021. **22**(1): p. 3-21.
89. King, A.D., et al., *Reversible Regulation of Promoter and Enhancer Histone Landscape by DNA Methylation in Mouse Embryonic Stem Cells.* Cell Rep, 2016. **17**(1): p. 289-302.

90. Pundhir, S., P. Poirazi, and J. Gorodkin, *Emerging applications of read profiles towards the functional annotation of the genome.* Front Genet, 2015. **6**: p. 188.
91. Smith, J., et al., *Promoter DNA Hypermethylation and Paradoxical Gene Activation.* Trends Cancer, 2020. **6**(5): p. 392-406.
92. Hafezi, F. and D. Perez Bercoff, *The Solo Play of TERT Promoter Mutations.* Cells, 2020. **9**(3).
93. Min, J. and J.W. Shay, *TERT Promoter Mutations Enhance Telomerase Activation by Long-Range Chromatin Interactions.* Cancer Discov, 2016. **6**(11): p. 1212-1214.
94. Chandra, V., et al., *Promoter-interacting expression quantitative trait loci are enriched for functional genetic variants.* Nat Genet, 2021. **53**(1): p. 110-119.
95. Morizono, A., et al., *Loss of BRCA1 expression and morphological features associated with BRCA1 promoter methylation status in triple-negative breast cancer.* J Hum Genet, 2021.
96. Tuupanen, S., et al., *The common colorectal cancer predisposition SNP rs6983267 at chromosome 8q24 confers potential to enhanced Wnt signaling.* Nat Genet, 2009. **41**(8): p. 885-90.
97. Wang, M.Y., et al., *Recurrent mutations at C-reactive protein gene promoter SNP position -286 in human cancers.* Cell Res, 2014. **24**(4): p. 505-8.
98. Ye, D., et al., *A novel SNP in promoter region of RP11-3N2.1 is associated with reduced risk of colorectal cancer.* J Hum Genet, 2018. **63**(1): p. 47-54.
99. Hu, N., et al., *Genome-wide association study of gastric adenocarcinoma in Asia: a comparison of associations between cardia and non-cardia tumours.* Gut, 2016. **65**(10): p. 1611-8.
100. Qiu, L.X., et al., *PSCA polymorphisms and gastric cancer susceptibility in an eastern Chinese population.* Oncotarget, 2016. **7**(8): p. 9420-8.
101. Wang, Y., et al., *SNP rs17079281 decreases lung cancer risk through creating an YY1-binding site to suppress DCBLD1 expression.* Oncogene, 2020. **39**(20): p. 4092-4102.
102. Aref-Eshghi, E., et al., *Genetic and epigenetic profiling of BRCA1/2 in ovarian tumors reveals additive diagnostic yield and evidence of a genomic BRCA1/2 DNA methylation signature.* J Hum Genet, 2020. **65**(10): p. 865-873.
103. Hu, X., et al., *Evolution of DNA methylome from precancerous lesions to invasive lung adenocarcinomas.* Nat Commun, 2021. **12**(1): p. 687.
104. Greenwald, W.W., et al., *Subtle changes in chromatin loop contact propensity are associated with differential gene regulation and expression.* Nat Commun, 2019. **10**(1): p. 1054.
105. Grubert, F., et al., *Landscape of cohesin-mediated chromatin loops in the human genome.* Nature, 2020. **583**(7818): p. 737-743.
106. Sanyal, A., et al., *The long-range interaction landscape of gene promoters.* Nature, 2012. **489**(7414): p. 109-13.
107. Zheng, H. and W. Xie, *The role of 3D genome organization in development and cell differentiation.* Nat Rev Mol Cell Biol, 2019. **20**(9): p. 535-550.
108. Ni, Z., et al., *TGR5-HNF4alpha axis contributes to bile acid-induced gastric intestinal metaplasia markers expression.* Cell Death Discov, 2020. **6**: p. 56.
109. Yu, J., et al., *CDX2 inhibits the proliferation and tumor formation of colon cancer cells by suppressing Wnt/beta-catenin signaling via transactivation of GSK-3beta and Axin2 expression.* Cell Death Dis, 2019. **10**(1): p. 26.
110. Nguyen, M.L., et al., *Transcriptional Enhancers in the Regulation of T Cell Differentiation.* Front Immunol, 2015. **6**: p. 462.
111. Kubo, N., et al., *Promoter-proximal CTCF binding promotes distal enhancer-dependent gene activation.* Nat Struct Mol Biol, 2021. **28**(2): p. 152-161.

112. Li, W., D. Notani, and M.G. Rosenfeld, *Enhancers as non-coding RNA transcription units: recent insights and future perspectives.* Nat Rev Genet, 2016. **17**(4): p. 207-23.
113. Sartorelli, V. and S.M. Lauberth, *Enhancer RNAs are an important regulatory layer of the epigenome.* Nat Struct Mol Biol, 2020. **27**(6): p. 521-528.
114. Zhang, Z., et al., *Transcriptional landscape and clinical utility of enhancer RNAs for eRNA-targeted therapy in cancer.* Nat Commun, 2019. **10**(1): p. 4562.
115. Ko, J.Y., S. Oh, and K.H. Yoo, *Functional Enhancers As Master Regulators of Tissue-Specific Gene Regulation and Cancer Development.* Mol Cells, 2017. **40**(3): p. 169-177.
116. Walker, B.A., et al., *Translocations at 8q24 juxtapose MYC with genes that harbor superenhancers resulting in overexpression and poor prognosis in myeloma patients.* Blood Cancer J, 2014. **4**: p. e191.
117. Krijger, P.H. and W. de Laat, *Regulation of disease-associated gene expression in the 3D genome.* Nat Rev Mol Cell Biol, 2016. **17**(12): p. 771-782.
118. Yan, T., et al., *Risk SNP-induced lncRNA-SLCC1 drives colorectal cancer through activating glycolysis signaling.* Signal Transduct Target Ther, 2021. **6**(1): p. 70.
119. Grampp, S., et al., *Genetic variation at the 8q24.21 renal cancer susceptibility locus affects HIF binding to a MYC enhancer.* Nat Commun, 2016. **7**: p. 13183.
120. Doni Jayavelu, N., et al., *Candidate silencer elements for the human and mouse genomes.* Nat Commun, 2020. **11**(1): p. 1061.
121. Cai, Y., et al., *H3K27me3-rich genomic regions can function as silencers to repress gene expression via chromatin interactions.* Nat Commun, 2021. **12**(1): p. 719.
122. Grotsch, B., et al., *The AP-1 transcription factor Fra1 inhibits follicular B cell differentiation into plasma cells.* J Exp Med, 2014. **211**(11): p. 2199-212.
123. Timblin, G.A. and M.S. Schlissel, *Ebf1 and c-Myb repress rag transcription downstream of Stat5 during early B cell development.* J Immunol, 2013. **191**(9): p. 4676-87.
124. Tsukumo, S., et al., *Bach2 maintains T cells in a naive state by suppressing effector memory-related genes.* Proc Natl Acad Sci U S A, 2013. **110**(26): p. 10735-40.
125. Painter, J.N., et al., *A Common Variant at the 14q32 Endometrial Cancer Risk Locus Activates AKT1 through YY1 Binding.* Am J Hum Genet, 2016. **98**(6): p. 1159-1169.
126. Gaszner, M. and G. Felsenfeld, *Insulators: exploiting transcriptional and epigenetic mechanisms.* Nat Rev Genet, 2006. **7**(9): p. 703-13.
127. Helmsauer, K., et al., *Enhancer hijacking determines extrachromosomal circular MYCN amplicon architecture in neuroblastoma.* Nat Commun, 2020. **11**(1): p. 5823.
128. Anastasiadou, E., L.S. Jacob, and F.J. Slack, *Non-coding RNA networks in cancer.* Nat Rev Cancer, 2018. **18**(1): p. 5-18.
129. Esteller, M., *Non-coding RNAs in human disease.* Nat Rev Genet, 2011. **12**(12): p. 861-74.
130. Guan, D., et al., *Switching cell fate, ncRNAs coming to play.* Cell Death Dis, 2013. **4**: p. e464.
131. Liu, L., et al., *Noncoding RNAs: the shot callers in tumor immune escape.* Signal Transduct Target Ther, 2020. **5**(1): p. 102.
132. Schwarzer, A., et al., *The non-coding RNA landscape of human hematopoiesis and leukemia.* Nat Commun, 2017. **8**(1): p. 218.
133. Mitra, S.A., A.P. Mitra, and T.J. Triche, *A central role for long non-coding RNA in cancer.* Front Genet, 2012. **3**: p. 17.
134. Dai, X., A.C. Kaushik, and J. Zhang, *The Emerging Role of Major Regulatory RNAs in Cancer Control.* Front Oncol, 2019. **9**: p. 920.
135. Shen, C., et al., *Long non-coding RNAs: Emerging regulators for chemo/immunotherapy resistance in cancer stem cells.* Cancer Lett, 2021. **500**: p. 244-252.

136. Li, P.F., et al., *Non-coding RNAs and gastric cancer.* World J Gastroenterol, 2014. **20**(18): p. 5411-9.
137. Romero-Cordoba, S.L., et al., *miRNA biogenesis: biological impact in the development of cancer.* Cancer Biol Ther, 2014. **15**(11): p. 1444-55.
138. Di Leva, G., M. Garofalo, and C.M. Croce, *MicroRNAs in cancer.* Annu Rev Pathol, 2014. **9**: p. 287-314.
139. Wadhwa, R., et al., *Gastric cancer-molecular and clinical dimensions.* Nat Rev Clin Oncol, 2013. **10**(11): p. 643-55.
140. Esquela-Kerscher, A. and F.J. Slack, *Oncomirs - microRNAs with a role in cancer.* Nat Rev Cancer, 2006. **6**(4): p. 259-69.
141. Yu, W., et al., *MicroRNA-195: a review of its role in cancers.* Onco Targets Ther, 2018. **11**: p. 7109-7123.
142. Jacinta-Fernandes, A., et al., *Allele-specific miRNA-binding analysis identifies candidate target genes for breast cancer risk.* NPJ Genom Med, 2020. **5**: p. 4.
143. Ryan, B.M., A.I. Robles, and C.C. Harris, *Genetic variation in microRNA networks: the implications for cancer research.* Nat Rev Cancer, 2010. **10**(6): p. 389-402.
144. Boni, V., et al., *Role of primary miRNA polymorphic variants in metastatic colon cancer patients treated with 5-fluorouracil and irinotecan.* Pharmacogenomics J, 2011. **11**(6): p. 429-36.
145. Xia, E., et al., *MicroRNA induction by copy number gain is associated with poor outcome in squamous cell carcinoma of the lung.* Sci Rep, 2018. **8**(1): p. 15363.
146. Spizzo, R., et al., *Long non-coding RNAs and cancer: a new frontier of translational research?* Oncogene, 2012. **31**(43): p. 4577-87.
147. Ma, L., et al., *LncRNAWiki: harnessing community knowledge in collaborative curation of human long non-coding RNAs.* Nucleic Acids Res, 2015. **43**(Database issue): p. D187-92.
148. Losko, M., J. Kotlinowski, and J. Jura, *Long Noncoding RNAs in Metabolic Syndrome Related Disorders.* Mediators Inflamm, 2016. **2016**: p. 5365209.
149. Prensner, J.R. and A.M. Chinnaiyan, *The emergence of lncRNAs in cancer biology.* Cancer Discov, 2011. **1**(5): p. 391-407.
150. Wang, Z. and Y. Zheng, *lncRNAs Regulate Innate Immune Responses and Their Roles in Macrophage Polarization.* Mediators Inflamm, 2018. **2018**: p. 8050956.
151. Zhao, Y., H. Sun, and H. Wang, *Long noncoding RNAs in DNA methylation: new players stepping into the old game.* Cell Biosci, 2016. **6**: p. 45.
152. Morris, K.V. and J.S. Mattick, *The rise of regulatory RNA.* Nat Rev Genet, 2014. **15**(6): p. 423-37.
153. Beckedorff, F.C., et al., *Long non-coding RNAs and their implications in cancer epigenetics.* Biosci Rep, 2013. **33**(4).
154. Sarkar, D., et al., *Epigenetic regulation in human melanoma: past and future.* Epigenetics, 2015. **10**(2): p. 103-21.
155. Dominguez, D., et al., *An extensive program of periodic alternative splicing linked to cell cycle progression.* Elife, 2016. **5**.
156. Peng, Z., C. Liu, and M. Wu, *New insights into long noncoding RNAs and their roles in glioma.* Mol Cancer, 2018. **17**(1): p. 61.
157. Peng, Z., C. Zhang, and C. Duan, *Functions and mechanisms of long noncoding RNAs in lung cancer.* Onco Targets Ther, 2016. **9**: p. 4411-24.
158. Yoshioka, H. and Y. Yoshiko, *The Roles of Long Non-Protein-Coding RNAs in Osteo-Adipogenic Lineage Commitment.* Int J Mol Sci, 2017. **18**(6).
159. Dey, B.K., K. Pfeifer, and A. Dutta, *The H19 long noncoding RNA gives rise to microRNAs miR-675-3p and miR-675-5p to promote skeletal muscle differentiation and regeneration.* Genes Dev, 2014. **28**(5): p. 491-501.

160. Sun, M. and W.L. Kraus, *From discovery to function: the expanding roles of long noncoding RNAs in physiology and disease.* Endocr Rev, 2015. **36**(1): p. 25-64.
161. Xie, S.S., et al., *Emerging roles of non-coding RNAs in gastric cancer: Pathogenesis and clinical implications.* World J Gastroenterol, 2016. **22**(3): p. 1213-23.
162. Delas, M.J. and G.J. Hannon, *lncRNAs in development and disease: from functions to mechanisms.* Open Biol, 2017. **7**(7).
163. Wu, P., et al., *Emerging role of tumor-related functional peptides encoded by lncRNA and circRNA.* Mol Cancer, 2020. **19**(1): p. 22.
164. Murugan, A.K., A.K. Munirajan, and A.S. Alzahrani, *Long noncoding RNAs: emerging players in thyroid cancer pathogenesis.* Endocr Relat Cancer, 2018. **25**(2): p. R59-R82.
165. Xue, Y., et al., *Association between lncrna PCGEM1 polymorphisms and prostate cancer risk.* Prostate Cancer Prostatic Dis, 2013. **16**(2): p. 139-44, S1.
166. Shen, C., et al., *Variant of SNP rs1317082 at CCSlnc362 (RP11-362K14.5) creates a binding site for miR-4658 and diminishes the susceptibility to CRC.* Cell Death Dis, 2018. **9**(12): p. 1177.
167. Abdi, E., et al., *SNP-SNP interactions of oncogenic long non-coding RNAs HOTAIR and HOTTIP on gastric cancer susceptibility.* Sci Rep, 2020. **10**(1): p. 16763.
168. Yang, Y., et al., *Recurrently deregulated lncRNAs in hepatocellular carcinoma.* Nat Commun, 2017. **8**: p. 14421.
169. Arun, G., S.D. Diermeier, and D.L. Spector, *Therapeutic Targeting of Long Non-Coding RNAs in Cancer.* Trends Mol Med, 2018. **24**(3): p. 257-277.
170. Li, J., et al., *A CRISPR Interference of CBP and p300 Selectively Induced Synthetic Lethality in Bladder Cancer Cells In Vitro.* Int J Biol Sci, 2019. **15**(6): p. 1276-1286.
171. Carter, P., *Site-directed mutagenesis.* Biochem J, 1986. **237**(1): p. 1-7.
172. Konieczny, A. and F.M. Ausubel, *A procedure for mapping Arabidopsis mutations using co-dominant ecotype-specific PCR-based markers.* Plant J, 1993. **4**(2): p. 403-10.
173. Fang, G., et al., *Discovering genetic interactions bridging pathways in genome-wide association studies.* Nat Commun, 2019. **10**(1): p. 4274.
174. Cowper-Sal lari, R., et al., *Breast cancer risk-associated SNPs modulate the affinity of chromatin for FOXA1 and alter gene expression.* Nat Genet, 2012. **44**(11): p. 1191-8.
175. Lee, B., et al., *Mapping genetic variability in mature miRNAs and miRNA binding sites in prostate cancer.* J Hum Genet, 2021.
176. Walavalkar, K., et al., *A rare variant of African ancestry activates 8q24 lncRNA hub by modulating cancer associated enhancer.* Nat Commun, 2020. **11**(1): p. 3598.
177. Nichols, C.A., et al., *Loss of heterozygosity of essential genes represents a widespread class of potential cancer vulnerabilities.* Nat Commun, 2020. **11**(1): p. 2517.
178. Lin, J., et al., *A miR-SNP biomarker linked to an increased lung cancer survival by miRNA-mediated down-regulation of FZD4 expression and Wnt signaling.* Sci Rep, 2017. **7**(1): p. 9029.
179. Huang, Q., et al., *A prostate cancer susceptibility allele at 6q22 increases RFX6 expression by modulating HOXB13 chromatin binding.* Nat Genet, 2014. **46**(2): p. 126-35.
180. Grillone, K., et al., *Non-coding RNAs in cancer: platforms and strategies for investigating the genomic "dark matter".* J Exp Clin Cancer Res, 2020. **39**(1): p. 117.
181. He, J., et al., *Fra-1 is upregulated in gastric cancer tissues and affects the PI3K/Akt and p53 signaling pathway in gastric cancer.* Int J Oncol, 2015. **47**(5): p. 1725-34.
182. Liu, W., et al., *A functional SNP rs1892901 in FOSL1 is associated with gastric cancer in Chinese population.* Sci Rep, 2017. **7**: p. 41737.
183. Wang, T., et al., *EGR1 is critical for gastrin-dependent upregulation of anion exchanger 2 in gastric cancer cells.* FEBS J, 2013. **280**(1): p. 174-83.

184. Luo, G.Q., et al., *Activator protein-1 involvement in proliferation inhibition by gene silencing of Twist in gastric cancer cells.* Pathology, 2011. **43**(7): p. 697-701.
185. Kim, M.S., et al., *Frameshift mutations of tumor suppressor gene EP300 in gastric and colorectal cancers with high microsatellite instability.* Hum Pathol, 2013. **44**(10): p. 2064-70.
186. Kim, M.S., et al., *Genetic and expressional alterations of CHD genes in gastric and colorectal cancers.* Histopathology, 2011. **58**(5): p. 660-8.
187. Zanke, B.W., et al., *Genome-wide association scan identifies a colorectal cancer susceptibility locus on chromosome 8q24.* Nat Genet, 2007. **39**(8): p. 989-94.
188. Shen, L., et al., *Clinical significance of POU5F1P1 rs10505477 polymorphism in Chinese gastric cancer patients receving cisplatin-based chemotherapy after surgical resection.* Int J Mol Sci, 2014. **15**(7): p. 12764-77.
189. Silberg, D.G., et al., *Cdx1 and cdx2 expression during intestinal development.* Gastroenterology, 2000. **119**(4): p. 961-71.
190. Garrison, W.D., et al., *Hepatocyte nuclear factor 4alpha is essential for embryonic development of the mouse colon.* Gastroenterology, 2006. **130**(4): p. 1207-20.
191. Ooi, W.F., et al., *Epigenomic profiling of primary gastric adenocarcinoma reveals super-enhancer heterogeneity.* Nat Commun, 2016. **7**: p. 12983.
192. Liz, J. and M. Esteller, *lncRNAs and microRNAs with a role in cancer development.* Biochim Biophys Acta, 2016. **1859**(1): p. 169-76.
193. Kim, K., et al., *HOTAIR is a negative prognostic factor and exhibits pro-oncogenic activity in pancreatic cancer.* Oncogene, 2013. **32**(13): p. 1616-25.
194. Niinuma, T., et al., *Upregulation of miR-196a and HOTAIR drive malignant character in gastrointestinal stromal tumors.* Cancer Res, 2012. **72**(5): p. 1126-36.
195. Gupta, R.A., et al., *Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis.* Nature, 2010. **464**(7291): p. 1071-6.
196. Pan, W., et al., *A functional lncRNA HOTAIR genetic variant contributes to gastric cancer susceptibility.* Mol Carcinog, 2016. **55**(1): p. 90-6.
197. Zhang, X., et al., *The identification of an ESCC susceptibility SNP rs920778 that regulates the expression of lncRNA HOTAIR via a novel intronic enhancer.* Carcinogenesis, 2014. **35**(9): p. 2062-7.
198. Hajjari, M. and S. Rahnama, *Association Between SNPs of Long Non-coding RNA HOTAIR and Risk of Different Cancers.* Front Genet, 2019. **10**: p. 113.
199. Bonaccorsi, L., et al., *Androgen receptor regulation of the seladin-1/DHCR24 gene: altered expression in prostate cancer.* Lab Invest, 2008. **88**(10): p. 1049-56.
200. Guo, H., et al., *Modulation of long noncoding RNAs by risk SNPs underlying genetic predispositions to prostate cancer.* Nat Genet, 2016. **48**(10): p. 1142-50.
201. Prensner, J.R., et al., *Transcriptome sequencing across a prostate cancer cohort identifies PCAT-1, an unannotated lincRNA implicated in disease progression.* Nat Biotechnol, 2011. **29**(8): p. 742-9.
202. Shang, Z., et al., *LncRNA PCAT1 activates AKT and NF-kappaB signaling in castration-resistant prostate cancer by regulating the PHLPP/FKBP51/IKKalpha complex.* Nucleic Acids Res, 2019. **47**(8): p. 4211-4225.
203. Song, Y.H., et al., *The important role of glycine N-methyltransferase in the carcinogenesis and progression of prostate cancer.* Mod Pathol, 2011. **24**(9): p. 1272-80.
204. Bi, M., et al., *Long non-coding RNA PCAT-1 over-expression promotes proliferation and metastasis in gastric cancer cells through regulating CDKN1A.* Gene, 2017. **626**: p. 337-343.
205. Grossi, E., et al., *A lncRNA-SWI/SNF complex crosstalk controls transcriptional activation at specific promoter regions.* Nat Commun, 2020. **11**(1): p. 936.

206. Qu, F., et al., *Genetic polymorphisms in circadian negative feedback regulation genes predict overall survival and response to chemotherapy in gastric cancer patients.* Sci Rep, 2016. **6**: p. 22424.
207. Sud, A., B. Kinnersley, and R.S. Houlston, *Genome-wide association studies of cancer: current insights and future perspectives.* Nat Rev Cancer, 2017. **17**(11): p. 692-704.